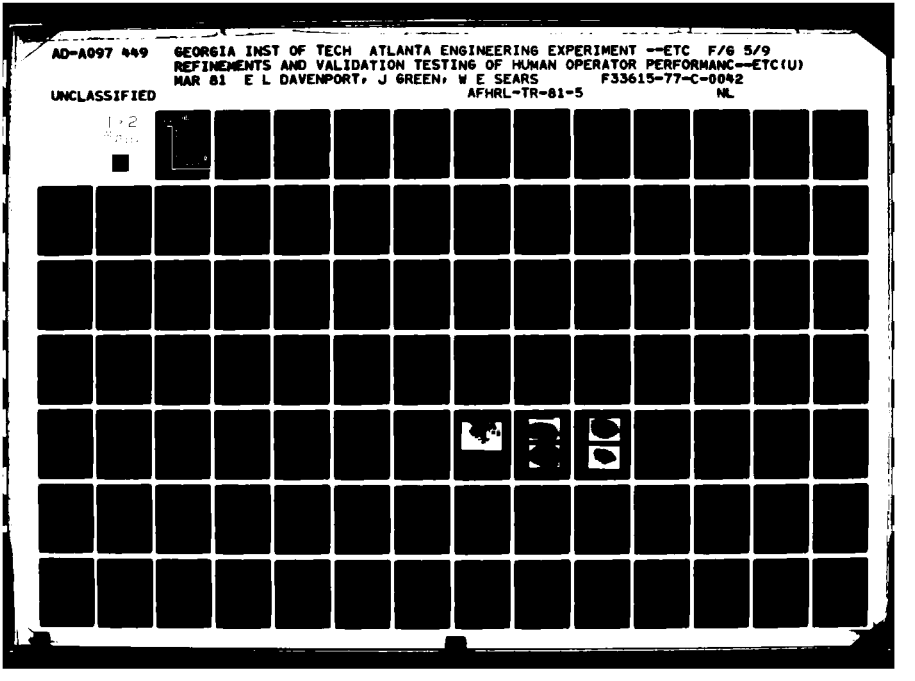
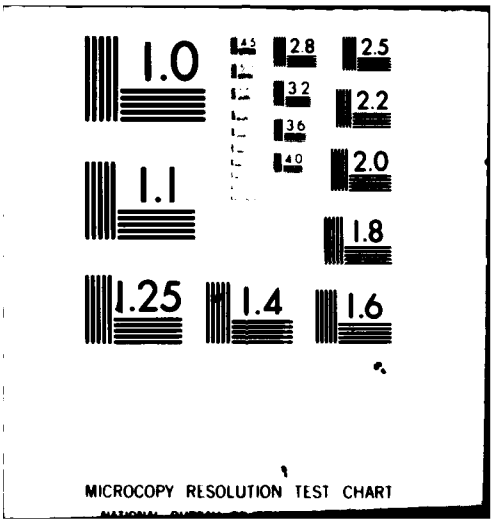
AD-A097 449 GEORGIA INST OF TECH ATLANTA ENGINEERING EXPERIMENT --ETC F/G 5/9
REFINEMENTS AND VALIDATION TESTING OF HUMAN OPERATOR PERFORMANC--ETC(U)
MAR 81 E L DAVENPORT, J GREEN, W E SEARS F33615-77-C-0042
UNCLASSIFIED AFHRL-TR-81-5 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART

AFHRL-TR-81-5

LEVEL II

(12)

AIR FORCE HUMAN RESOURCES LABORATORY

AD A097449

DTIC FILE COPY

REFINEMENTS AND VALIDATION TESTING OF HUMAN OPERATOR PERFORMANCE EMULATOR (HOPE)

By

Esther Lee Davenport
Joanne Green
William E. Sears, III
Harold F. Engler
Georgia Institute of Technology
Engineering Experiment Station
Atlanta, Georgia 30332

OPERATIONS TRAINING DIVISION
Williams Air Force Base, Arizona 85224

March 1981

Final Report

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When U.S. Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This final report was submitted by the Georgia Institute of Technology, Engineering Experiment Station, Atlanta, Georgia 30332, under Contract F33615-77-C-0042, Project 2313, with the Operations Training Division, Air Force Human Resources Laboratory (AFSC), Williams Air Force Base, Arizona 85224. Thomas Longridge was the Contract Monitor for the Laboratory.

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

MARTY R. ROCKWAY, Technical Director
Operations Training Division

RONALD W. TERRY, Colonel, USAF
Commander

AD-A

DEPARTMENT OF THE AIR FORCE
AIR FORCE HUMAN RESOURCES LABORATORY (AFSC)
BROOKS AIR FORCE BASE, TEXAS 78235

REPLY TO ATTN OF: TSR

7 APR 1981

SUBJECT: Correction to DD Form 1473 - AFHRL-TR-81-5, "Refinements & Validation Testing of Human Operator Performance Emulator (HOPE)"

TO: Defense Technical Information Center
Cameron Station
Alexandria VA 22314

Please correct Block 10, Work Unit Number, of subject technical report to read 2313-T3-20. Questions may be directed to this office, AV 240-3877.

Nancy A. Perrigo

NANCY A. PERRIGO
Chief, STINFO Office

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AFHRL-TR-81-5	2. GOVT ACCESSION NO.: AD-A097 449
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): REFINEMENTS AND VALIDATION TESTING OF HUMAN OPERATOR PERFORMANCE EMULATOR (HOPE)	5. TYPE OF REPORT & PERIOD COVERED: Final
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Esther Lee Davenport, Harold F. Engler, Joanne Green, William E. Sears, III	8. CONTRACT OR GRANT NUMBER(s): F33615-77-C-0042
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Georgia Institute of Technology, Engineering Experiment Station, Atlanta, Georgia 30332	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F, 2313T520, T5
11. CONTROLLING OFFICE NAME AND ADDRESS: HQ Air Force Human Resources Laboratory (AFSC), Brooks Air Force Base, Texas 78235	12. REPORT DATE: March 1981; 13. NUMBER OF PAGES: 140
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Operations Training Division, Air Force Human Resources Laboratory, Williams Air Force Base, Arizona 85224	15. SECURITY CLASS. (of this report): Unclassified; 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

control strategy
simulation of motor behavior
strategy measurement
preview tracking
manual control learning
mathematical modeling

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

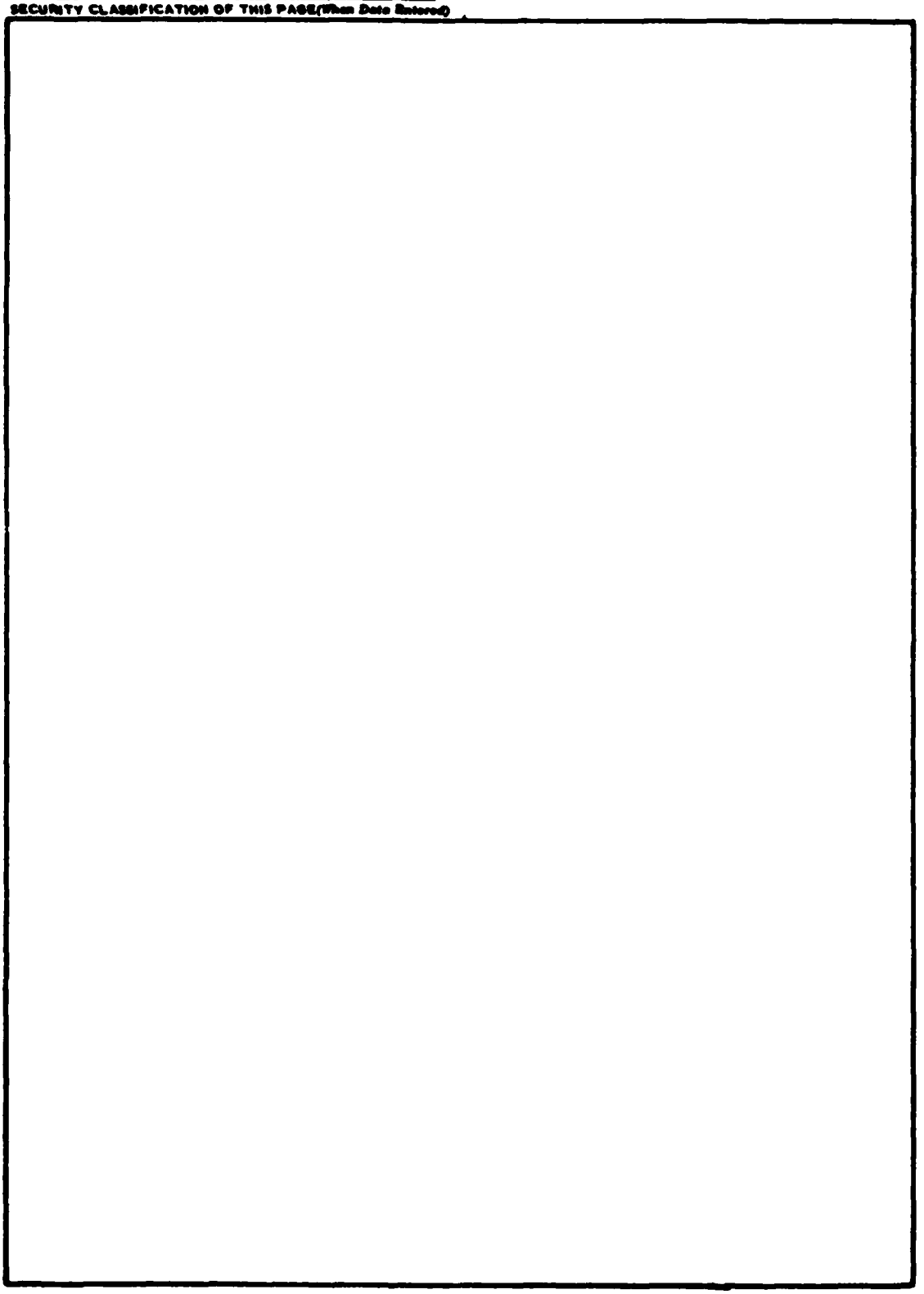
There is a need for an improved measurement system for use in flying training. The measurement system is needed to guide the design of effective flight training simulators, to aid in selection of the most effective training procedures, and to accurately specify when an individual trainee is ready to transfer from ground-based simulator to real flight.

The research reported here and in AFHRL-TR-79-60 was aimed at developing and initially testing a totally new approach to the needed measurement system.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified 153850

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)



PREFACE

This work was conducted under Air Force Human Resources Laboratory Project 2313-T5-20 (formerly 2313-T3-03), Contract F33615-77-C0042. Ms. Patricia A. Knoop, Advanced Systems Division, Air Force Human Resources Laboratory (AFHRL), served as the Program Manager for the Air Force during the major portion of the contract. Dr. Thomas Longridge of the Operations Training Division, AFHRL, served as Program Manager during the final 2 months. This report is the second of two reports which, taken together, describe the work over the 3-year contract period. The first report is AFHRL-TR-79-60, entitled Human Operator Control Strategy Model, April 1980.

The research was conducted at the Engineering Experiment Station of the Georgia Institute of Technology under the technical supervision of Esther Lee Davenport, Project Director. It was administered under the direct supervision of William E. Sears, III, Chief, Special Projects Division, Systems Engineering Laboratory, and the overall supervision of Robert P. Zimmer, Laboratory Director.

The authors are especially appreciative to Ms. Knoop for conceiving the approach taken here and for providing timely guidance.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Concluded)

LIST OF FIGURES

LIST OF FIGURES (Concluded)

LIST OF TABLES

LIST OF TABLES (Concluded)

SECTION I

INTRODUCTION

This report is the second of two major technical reports describing the development and testing of a computer simulation of manual control behavior. The simulation is called the Human Operator Performance Emulator (HOPE) and is unique in at least three respects. First, the simulation contains representations of structures and processes believed important in continuous control behavior, and thus has a degree of psychological validity not characteristic of other models of the same behavior. Second, the simulation models the learning of control behavior, and thus models behavior of both trained and untrained operators. Third, and probably most importantly, the simulation includes representation of a construct called control strategy, and was designed to measure control strategy in humans.

The previous report (see Engler, Davenport, Green, & Sears, 1980) described in detail the research problem being addressed, the theory of continuous control behavior behind the adopted research approach, the operation of HOPE, and preliminary tests of the HOPE simulation. The remainder of this section summarizes some of these ideas, with emphasis on the rationale for the research approach. For further detail, the reader is referred to Engler et al. (1980). The remaining sections of this report describe an extension of the work described in that earlier report.

Section II describes refinements made to HOPE, including descriptions of symptoms suggesting the need for refinement, details of the refinements made in HOPE processes, and the resulting outcomes of these changes. Section III describes changes that were made in the control strategy measurement procedure. The problems stimulating these changes and the outcomes of implementing the changes are described. HOPE has received additional validation testing, the purposes, methods, results and outcomes of which are described in Section IV. Finally, Section V provides an overall assessment of HOPE, its ability to measure human behavior, and the usefulness of the general research approach that has been applied. This final section also describes and prioritizes recommendations for further research, both to improve HOPE and to extend the present research approach to measurement of control strategy.

A. Summary of the Problem Being Addressed

Continuous manual control behavior is a fundamental component of a variety of important skills. These skills include driving a car or flying an airplane, where a smooth sequence of accurate manual movements must be executed in response to a presented pattern such as a road or the demand for a particular flight maneuver.

A variety of measures have been used to describe the quality of control behavior. These include measures such as average absolute error, root mean square error, or time on target. Although they have been usefully applied, these measures are problematic in a number of

ways. As demonstrated by Obermayer, Swartz and Muckler (1962), they are sometimes inconsistent in their response to experimental manipulations. Several of the commonly used measures are not Gaussian in their distribution and therefore cannot be analyzed through use of parametric statistics (Poulton, 1974). Finally, and most importantly, these measures are not adequately sensitive to changes and differences in behavior in a variety of important situations.

One of the areas where more sensitive measures are needed is flight simulation research. Current measures cannot detect the differential effects of the variety of cues which might be incorporated in simulators. As Knoop (1978) points out, existing measures "do not have the necessary characteristics to support the type of flight simulation research that entails accounting for the perception and utilization of cues." Measures reflecting such differences could be used to identify the subset of cues which actually affect learning and performance in real flight. These cues could then be included in the flight simulators used in training, greatly increasing cost-effectiveness in simulator design.

More sensitive measures are also needed to assess individual behavior. Current measures do not detect the subtle differences between individuals which reflect differences in skill. It is these differences which may be predictive of the transferability of training to real flight, or of success in difficult real flight conditions (e.g., engine failure, bad weather).

The research reported here and in the earlier technical report was aimed at developing a more sensitive measure of manual control behavior which might be useful for:

- describing the effects on behavior of different cues,
- better describing individual differences in skill both during and after training, and
- predicting trainee readiness for real flight.

B. Research Approach

There are three major foci of the research approach described in the two technical reports, Engler et al. (1980) and this one. The first distinctive focus is on control strategy as an important, measurable, trainable aspect of human behavior. The second focus is the use of a computer simulation of human mental processes to measure control strategy. The third important focus is on the validation of the concept of control strategy and its measurement by use of a simulation. These three foci are briefly described in the following discussion.

1. Control Strategy

Currently available measures focus on describing aspects of behavior associated with system error (e.g., position error). The research approach adopted here focused on measuring aspects of behavior which comprise "control strategy." Many researchers have argued that "strategy" is an important aspect of a variety of behaviors (see Moray, 1975; Welford, 1968; Alegria, 1975). The present research approach assumes

that control strategy is an important determinant of the style or pattern of control behavior, and that its measurement would describe the subtle differences in behavior for which measures are needed.

Control strategy was defined as a set of parameter values which determine the functioning of mental processes important in continuous manual control. The parameters can be divided into three categories:

- criteria for control behavior and performance,
- stimulus cues on which to base performance, and
- the sequence for mental decision-making processes.

2. The Human Operator Performance Emulator (HOPE)

It was decided to measure control strategy in humans through use of a psychologically-based computer model of continuous control behavior containing a representation of control strategy. The simulation, called the Human Operator Performance Emulator (HOPE), includes representations of basic psychological processes and structures (e.g., perception, memory, performance monitoring) and models continuous control behavior both during learning and after learning asymptote. The type of continuous control behavior which is simulated is preview tracking. Control strategy is represented in HOPE by three control strategy parameters (CSPs) which determine the operation of other processes in the simulation. The three parameters are Command Operative Time (COT), which functions as an upper limit on the frequency of control movements; ERRLIM, an internal performance standard which determines whether performance is judged acceptable or not; and ADJUST, which determines, especially in learning, the magnitude of response to excessive error and the compensation for lag or gain in the control dynamics.

HOPE measures human preview tracking behavior as follows. Given a numerical representation of the track and values for COT, ERRLIM, and ADJUST, HOPE predicts the control stick positions that a human would use to align a cursor with the center of a moving track. HOPE is run multiple times with different sets of CSP values, each time producing a "model" of control strategy-guided behavior. The human behavior to be measured in a given time bin is compared to the behaviors in that time bin of the different HOPE models. The CSP values modulating the HOPE model which best match the human behavior are used to infer the nature of human control strategy in that time bin.

HOPE has been explicitly designed for measuring individual differences in cue utilization and control style. These are believed to be reflected in the aspects of control strategy defined and measured by HOPE. The research approach represented by HOPE has a variety of other advantages over other current attempts to use models for measurement purposes. In contrast to the optimal control model (Kleinman, Baron and Levison, 1970) HOPE models the learning of control behavior and thus can be applied to measurement of both trained and untrained behavior. It can model control over vehicles which have linear or non-linear control dynamics. Finally, its representation of psychological processes and structures provides a clearer idea of the functioning of the human mind and of the effects of control strategy on human infor-

mation processing than is available from non-psychologically based models.

3. Validation

Validation activities have received major emphasis throughout the research. First, there has been an attempt to define control strategy so as to permit inclusion of human characteristics, such as perceptual learning abilities, which have already been demonstrated in the psychological literature. Second, the theory on which HOPE is based includes fundamental psychological principles, such as that of limited processing capacity, which have been repeatedly demonstrated as important in skill learning. Refinements made to HOPE and described in this report have been guided by the necessity to remain consistent with these fundamental principles. Third, two experiments have been conducted in order to begin to collect the data on which the judgement of validity must ultimately be based (see Section V of Engler et al., 1980, and Section IV of this report). In these two experiments, it has been possible to examine both the degree to which the measures taken seem to be consistent with control strategy as it was defined (i.e., construct validity) and the extent to which measures of control strategy can be used to predict other measures, either of control strategy or of performance in the tracking task used in the testing (i.e., criterion-related validity).

C. Summary of Previous Work on this Contract

The work reported here was preceded by other critical theoretical and experimental work. The tasks that were completed prior to those reported here are listed and briefly described below. Detailed discussion is contained in Engler et al. (1980).

1. Development of a Theory of Manual Control Learning and Performance

This theory forms the basis for the HOPE simulation. It enumerates mental processes believed important to continuous manual control learning and performance, and divides these into two basic categories, automatic and decision-making processes. The theory defines control strategy and its relation to mental processing and learning. Control strategy is believed to become task-specific over the course of learning a new control task. It is hypothesized that information organized in a variety of mental structures is involved in continuous control behavior. These organizations of information include a task controller model which stores associations between control positions and vehicle states, an input model which can be used to predict upcoming stimuli, and a neuromuscular model which represents limits in control behavior due to neuromuscular processing time. Finally, the theory attributes control learning to two major factors--the development of task-specific control strategies, and the accumulation of accurate information in the internal models.

2. Development of the HOPE Simulation

The HOPE simulation reflects the theory of control learning and performance and models a subset of the structures, processes and parameters hypothesized in the theory. HOPE simulates a particular type of continuous

control behavior--preview tracking. Basic elements of the HOPE simulation are described in Section II of this report.

3. Preliminary Validation Testing of HOPE

HOPE received preliminary testing in a laboratory experiment designed to test its ability to match and measure human behavior. Three research questions were addressed: Can HOPE match human behavior to a significant extent? Does control strategy as identified by HOPE vary over the course of learning? Does control strategy as identified by HOPE reflect differences between training conditions?

Human subjects were tested in one of four conditions of preview tracking, and their control behavior was recorded. HOPE models of behavior reflecting different control strategies were used to infer human control strategy at different points in learning and in different conditions of tracking. The results of the preliminary testing suggested that HOPE did match human behavior to an acceptable extent, and that HOPE could identify changes in human control strategy. The results also revealed a variety of problematic symptoms in HOPE behavior and in measurement procedures which were addressed in the work described in this report (see Sections II and III).

SECTION II

REFINEMENTS TO HOPE

A. Introduction

HOPE is a psychologically based computer simulation of continuous manual control which includes a representation of control strategy. HOPE currently simulates preview tracking behavior based on use of visual information only. It generates control stick positions (commands) which might be used to keep an externally viewed cursor on the center of a track. However, HOPE was developed from a broader theory of psychomotor behavior which applies to a wider range of behaviors and information processing activities (see Engler et al., 1980, Section III).

This chapter begins with a summary of the structure and operation of HOPE. Details are provided in Section IV of Engler et al. (1980). The major part of the chapter then describes refinements made to HOPE based on data collected during preliminary tests of the model.

B. Summary of HOPE Structure and Operation

Figure 1 indicates the basic processes and structures of HOPE. HOPE is a hierarchical model in which there is a clear distinction made between two levels of processes. Although the psychological literature has variously labeled these levels as conscious-subconscious, attentive-preattentive (Neisser, 1967), controlled-automatic, most models of information processing recognize a distinction between processes which demand attention and must be performed one at a time and processes which do not demand attention and can be carried out in parallel. In HOPE, one level of processing is represented by a Supervisory Processor that can perform a variety of operations but only in a serial fashion. The constraint that the Supervisory Processor can perform only one function at a time links HOPE with single-channel models of human information processing (Welford, 1952). In contrast, the second level of processing includes a number of lower-level subsidiary processors, each dedicated to a single process, but which can operate in parallel. Assignment of a process as a Supervisory Processor function or as a subsidiary function is made in accordance with descriptions of the relative demands of different mental operations on the human's limited processing capacity (Kahneman, 1973; Posner and Boies, 1971). Supervisory Processor functions are those demanding more processing capacity.

The following discussion describes the basic function and operation of HOPE processes and associated structures.

1. Subsidiary Processes and Associated Structures

a. Perception process--The Perception Process acquires external information necessary for task performance and translates it into a form usable by other HOPE processes. The Perception Process provides to other processes information about the current cursor position

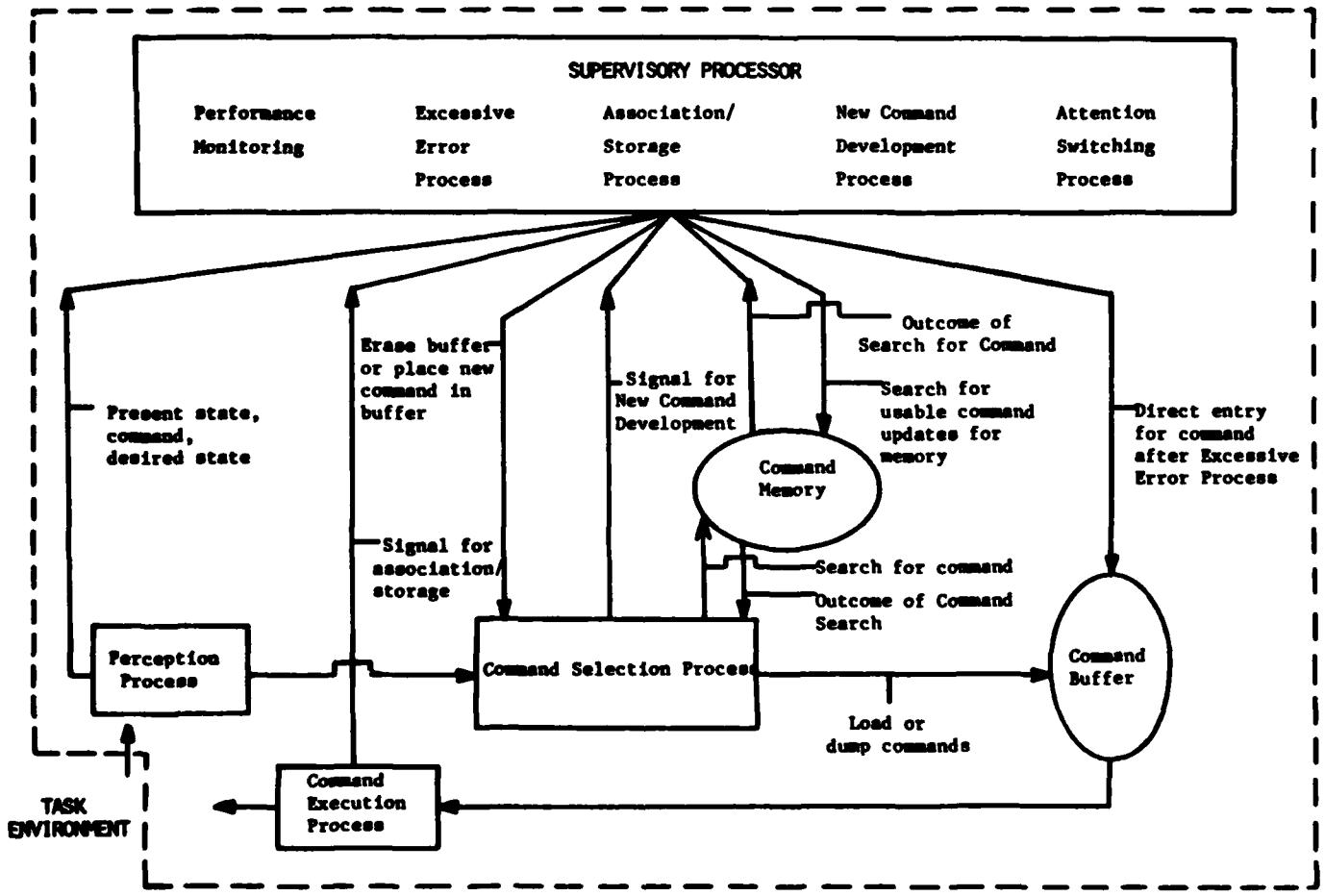


Figure 1. Human Operator Performance Emulator (HOPE)

(current external state), the desired cursor position (desired external state, usually an upcoming track point), and the current control stick position (command). The intermittency of human perception (Bertelson, 1967) is represented by the fact that the Perception Process makes information available only at discrete intervals--i.e., every 40 msec.

b. Command memory--The Command Memory is the HOPE representation of long-term memory. It is a two-dimensional memory array containing commands associated with specified transitions between pairs of cursor positions, or external states. It is organized in an associationist fashion whereby the commands are located in memory at a point addressed by the cursor positions that preceded and followed the command. The commands which are stored at each address depend, in part, on Command Operative Time (COT), one of the control strategy parameters. The stored command will accomplish in one COT the state transition described by its address. The command memories of HOPE models associated with different COTs will have different commands stored at the same address. HOPE begins a new task with a blank Command Memory; the memory becomes loaded with commands with experience in the task. Prior experience would be represented by a partially filled Command Memory.

c. Command selection processs--This process performs the function of looking up commands in the Command Memory, and of updating identification of the external states. The Command Selection Process places located commands in the Command Buffer, in which a backlog of commands is often developed because Command Execution usually occurs more slowly than Command Selection. For this reason, the Command Selection Process is frequently selecting commands for use in the future, rather than for use as the next command to be executed. The Command Selection Process locates commands in the Command Memory by addressing it according to the 'desired state' (an upcoming track point) and the 'last predicted external state' (the cursor position predicted to result from the last selected command).

The Command Selection Process also updates the state labels so that the desired external state associated with the located command is identified as the next 'last predicted external state'. Information from the Perception Process is used to identify the next desired external state.

d. Command buffer--This is the HOPE representation of human short-term memory. The Command Buffer stores located or generated commands (see upcoming discussion of Satisfactory Command Search and Excessive Error Process) until they can be executed. Since commands are produced more frequently than they can be executed (see Command Execution Process), the Command Buffer frequently stores a backlog of commands, in the order in which they were produced. While the Command Memory is theoretically unlimited in size, like human long-term memory, the Command Buffer stores only a limited number of commands, similar to human short-term memory.

e. Command execution process--This process is responsible for applying the commands stored in the Command Buffer. The Command Execution Process takes the command at the top of the buffer and executes

it for a small amount of time. The command duration depends primarily on one of the control strategy parameters, Command Operative Time, and can vary between 40 and 240 msec in different HOPE models of strategy-modulated behavior. When the command duration expires, the Command Execution Process requests the next command from the buffer. If none is available, the previous command is repeated.

2. Supervisory Processor Functions

The remaining processes cannot be performed in parallel, in contrast to the subsidiary processes just discussed. The serial operation of Supervisory Processor functions is organized by means of both a queuing of requests for Supervisory Processor attention and an interrupt procedure. In general, requests for Stimulus-Response Association, Satisfactory Command Search and Attention Reallocation are recorded in a queue, and are serviced by the Supervisory Processor in turn. Requests for the Excessive Error Process interrupt this queuing, and are served immediately. Performance Monitoring alternates with the other Supervisory Processor functions.

a. Stimulus-response association--This process updates the contents of the Command Memory. Each time Command Execution begins, a request for a Stimulus-Response Association is sent to the Supervisory Processor. When the Supervisory Processor services this request, the Stimulus-Response Association Process uses information from the Perception Process to place in the Command Memory a command addressed by a current external state and a desired state. Under ideal conditions, when the Supervisory Processor is immediately available to service a Stimulus-Response Association request, the information stored is highly accurate--that is, execution of the stored command will cause a state transition between the external states addressing its memory location. However, when servicing of the Stimulus-Response Association request is delayed, the information which is associated and stored may be somewhat inaccurate. Accurate information about external states associated with the command to be stored is no longer available from the Perception Process because it has stored more recent external state information during the course of the time delay. For this reason, the command stored in memory may not cause the exact state transition specified by the states addressing the memory location. To represent the effects of repeated encounters with the same transition, each new command to be stored in a Command Memory location is averaged with the average of other commands previously stored in that location.

b. Satisfactory command search--This process generates a command when the Command Selection Process is unable to locate a stored command in the Command Memory. It is especially important early in HOPE learning when the Command Memory is relatively empty. Since the original procedures for Satisfactory Command Search are quite complex and have received considerable refinement, they will not be discussed here. The refined procedures are discussed later in this section.

The command generated by the Satisfactory Command Search is placed in the Command Buffer.

c. Performance monitor--The Performance Monitor determines if performance is within acceptable error limits, as defined by ERRLIM, one of the HOPE control strategy parameters. The Performance Monitor compares the current and desired external states. If the difference is greater than the magnitude of ERRLIM, and if no corrective action has been recently initiated, then the Performance Monitor invokes the Excessive Error Process.

d. Excessive error process--The Excessive Error Process takes steps to avoid further error and generates a command with which generation of a new command string can begin. Since detection of an unacceptable error questions the quality of commands lined up in the Command Buffer, the buffer is dumped. Also, any plans for Supervisory Processor function may be incorrect, so the Supervisory Processor queue is dumped. (This aspect has received some refinement--see Section II-D.4).

As with the Satisfactory Command Search, the procedures for Excessive Error Process generation of a command are quite complex and have received considerable refinement, so they will be discussed later only with respect to the refined version. It should be noted here, however, that the magnitude of response to excessive error is strongly influenced by ADJUST, the third control strategy parameter in HOPE. Small values of ADJUST result in a less aggressive response than do large values.

The command generated by the Excessive Error Process is placed in the Command Buffer, and the predicted result is provided to the Command Selection Process to initiate reselection of commands.

e. Attention reallocation process--This process allows attention to be temporarily switched away from the tracking task. This can occur if there is a sufficient backlog of commands in the Command Buffer or when commands have been planned for all states out to the maximum preview available. It will not occur if there is a "best-guess" command in the buffer. This is a type of command generated by the Satisfactory Command Search or the Excessive Error Process, and its results are usually unpredictable (see Section II-D.3). Attention is switched back to tracking when there is only one command remaining in the Command Buffer, and Command Selection must continue.

Since the testing to date has not included tasks other than the tracking task, the Attention Reallocation Process is inoperative in the current version of HOPE.

3. Psychological Validity of HOPE

HOPE was designed to have a considerable degree of psychological validity. This is achieved by including representations of basic psychological processes believed to be important in continuous motor control. Although it may be impossible to truly specify and thus accurately represent the details of mental processes and structures (see Anderson, 1978), the fact that HOPE contains logically reasonable representations gives it a degree of psychological validity not approached in other models of continuous control behavior.

As previously discussed, HOPE reflects current theory on the amount of processing capacity required by different mental operations, as well as theories of memory which postulate the importance of both a short-term and long-term memory in behavior. The HOPE simulation addresses many of the criticisms aimed at mathematical models of control behavior, such as the optimal control model (Kleinman, Baron, & Levison, 1970).

Its Perception Process allows information to be input only at discrete intervals, thus reflecting the human operator characteristic of intermittency. The execution of each command in HOPE is delayed until the previous command terminates, causing the model to exhibit behavior similiar to that associated with the human psychological refractory period (Welford, 1968). The storage of commands in the Command Memory allows it to improve performance on the basis of previous experience, and thus, to "learn" over the course of performance in contrast with the optimal control model. The theory on which HOPE is based also includes a learned input.

Finally, HOPE is an imperfect performer, as, of course, are people. There are two major sources of error of HOPE. The first results from generation of best-guess commands by the Satisfactory Command Search or Excessive Error Process (see Section II-D.3), and represents the use of "educated guesses" by humans. The second source of error occurs when the Supervisory Processor cannot immediately service a request for a Stimulus-Response Association, and inaccurate information is stored in the Command Memory. This type of error represents one of the consequences of limited processing capacity in humans.

C. Symptoms Suggesting Need for Refinements

The initial version of HOPE received preliminary testing which is described in detail in Engler et al. (1980). In these tests, subjects tracked either a ¼ Hz maximum-frequency preview track or a ½ Hz maximum-frequency preview track. The results of the tests indicated that HOPE was able to match human control behavior well within a pre-determined criterion for acceptable matching. Furthermore, certain aspects of the human control strategy inferred through use of HOPE varied in a way consistent with current theory on control strategy variation as a function of learning and training conditions.

Analysis of data from the preliminary tests suggested a variety of symptoms indicating how HOPE might be further improved and refined. These symptoms and the refinements addressing them are described in the remainder of this section.

1. Data Analyses Performed in Symptom Identification

The symptoms suggesting the need for model refinements were the outcomes of a variety of analyses of the data collected during the initial testing of HOPE, and generated to document the operation of HOPE. The following data were examined:

- Plots of control stick positions, both those generated by selected humans and certain HOPE models. These allow

comparison of the styles (e.g., smoothness) of human and HOPE model control stick manipulation.

- Listings of control stick positions from selected operators and models. These allow comparison of the values of commands tried by humans and models to achieve specified cursor transitions.

- Plots of cursor positions produced by humans and by HOPE models. These allow comparison of the style of human and model cursor manipulation, and the kinds of position errors made by each.

- Root mean square (RMS) difference values (computed between humans and their best-fitting models) averaged within and across trials for each person. These provide a measure of the ability of best-fit models to match human behavior.

- Minimum RMS difference values for the first 12 time bins of the first trial of tracking for all operators, as selected by comparing first trial operator behaviors with model behaviors in comparable bins in all five trials. These values determine whether model behavior best matches human behavior at a comparable point in experience, or at an earlier or later point.

- Changes in best-fit model control strategy parameters (CSPs) as a function of conditions of testing and of practice. These values provide data for judging whether HOPE's estimates of human control strategy vary with training condition and learning.

- The commands valid for producing specified changes in cursor position, for the control dynamics used in the test. These values provide more information about the learning required for plant control, and the relation between the commands required to make similar state transitions.

- The value of commands stored in the HOPE Command Memory at various points in training, for selected models. These allow examination of the process of Command Memory development.

- Mean absolute position error of all human operators and their best-fit models, within and across trials. These allow comparison of human and HOPE model learning trends.

- Diagnostics generated during model execution, such as command string length, calls to the Command Selection Process, calls to the Excessive Error Process, etc., for all models. These document how the model is learning.

- Mean absolute position error for all 75 models, averaged within and across trials, and across conditions. These allow comparison of model learning trends and identification of which models learn most quickly.

2. Outcomes of Data Analysis

Examination of these data exposed a variety of symptoms which formed the basis for designing model refinements. These symptoms are named and described briefly in the paragraphs that follow. In the remainder of this section, the original version of HOPE will be referred to as HOPE1; the refined version will be referred to as HOPE2.

a. Differences between model and human control stick patterns--Plots of control stick positions for selected bins of human behavior and corresponding best-fit models indicated several differences between the style of human and model control stick manipulation. Human behavior was relatively smooth and continuous, while model behavior was not (see Figure 2).[1] Model behavior showed greater variability than that of humans; there was a greater fluctuation of control stick position within a given limit of time (see Figure 2 for illustration). One aspect of this was the appearance of "spikes" in model behavior--commands which were quite divergent from the preceding and following command sequence, producing noticeable upward and downward "spikes" in plots of model control stick positions. Such spikes and variability were particularly apparent when the model reversed the direction of tracking.

b. Poorer model matches for human behavior on early trials--The root mean square (RMS) difference value is a measure of the position difference between model and human behavior and is used to select the best-fit model for a given interval of time. The RMS difference values of best-fit models can be used to gauge the goodness of fits to human behavior. The smaller the RMS difference value, the better the fit. Figure 3 displays the average RMS difference values for subjects in each condition for each trial. The values were greater for best-fit models for early trials than for later trials of behavior. This trend was more consistent for subjects in ¼ Hz track conditions. It suggested that HOPE was not able to match equally well human behavior at different stages of learning. Ideally, HOPE should be able to match human behavior equally well at all points in experience.

c. Poorer model matches for ½ Hz track conditions--As can be seen in Figure 3, the RMS difference values of best-fit models were greater for ½ Hz than for ¼ Hz track behavior. Ideally, HOPE should be able to match behavior equally well in either track condition.

---

[1] The apparent "jitter" in human behavior (especially near control stick extremes) is an artifact of noise in the A/D conversion, and does not represent human control behavior.

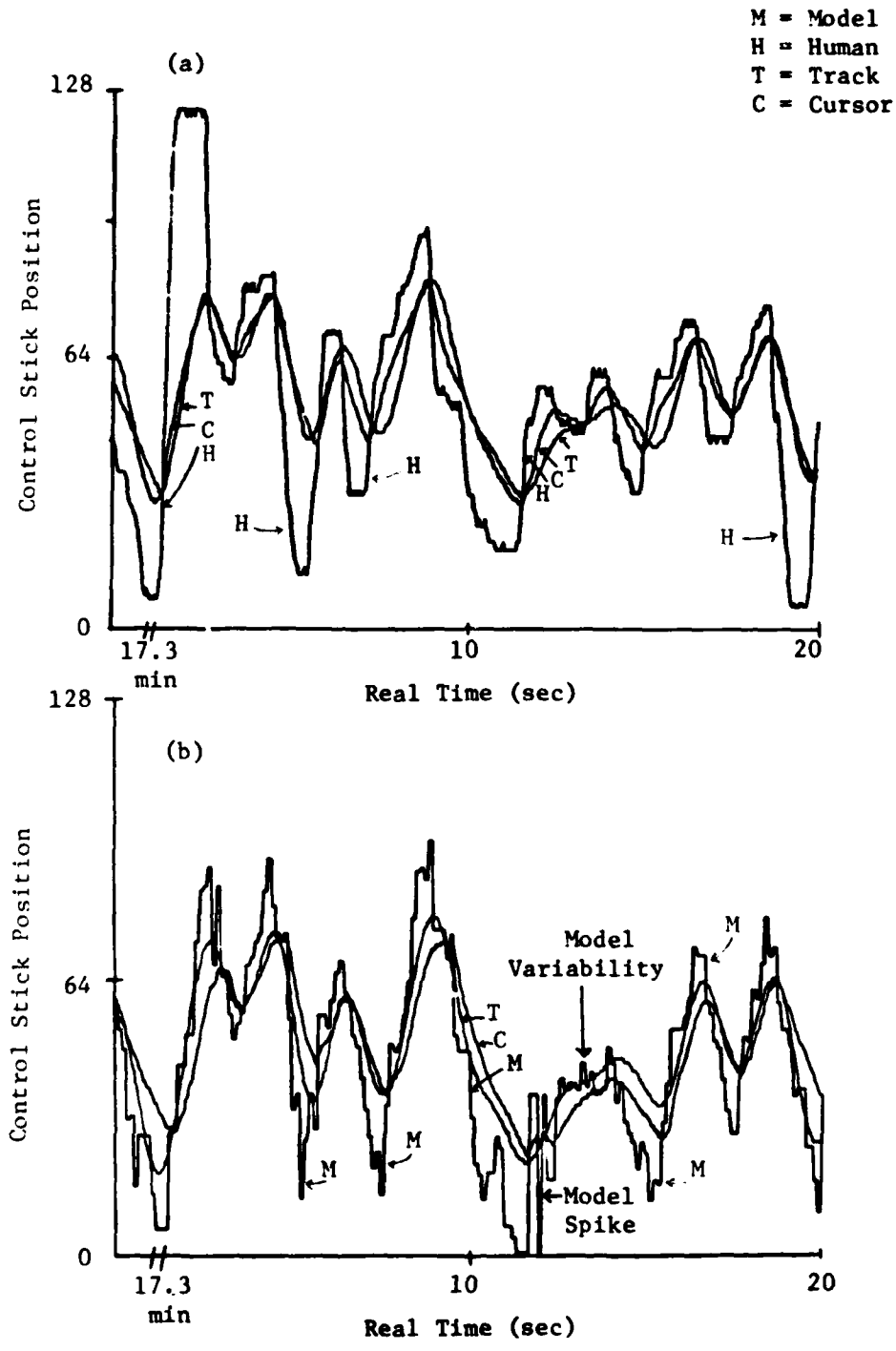


Figure 2. Control Stick Positions Used by a Human Subject (a) and a HOPE1 Model (b)--These positions were used during the same 20 sec time bin of tracking in a 1/2 Hz track condition in preliminary tests. Model behavior is considerably more spiky and variable than is human behavior.

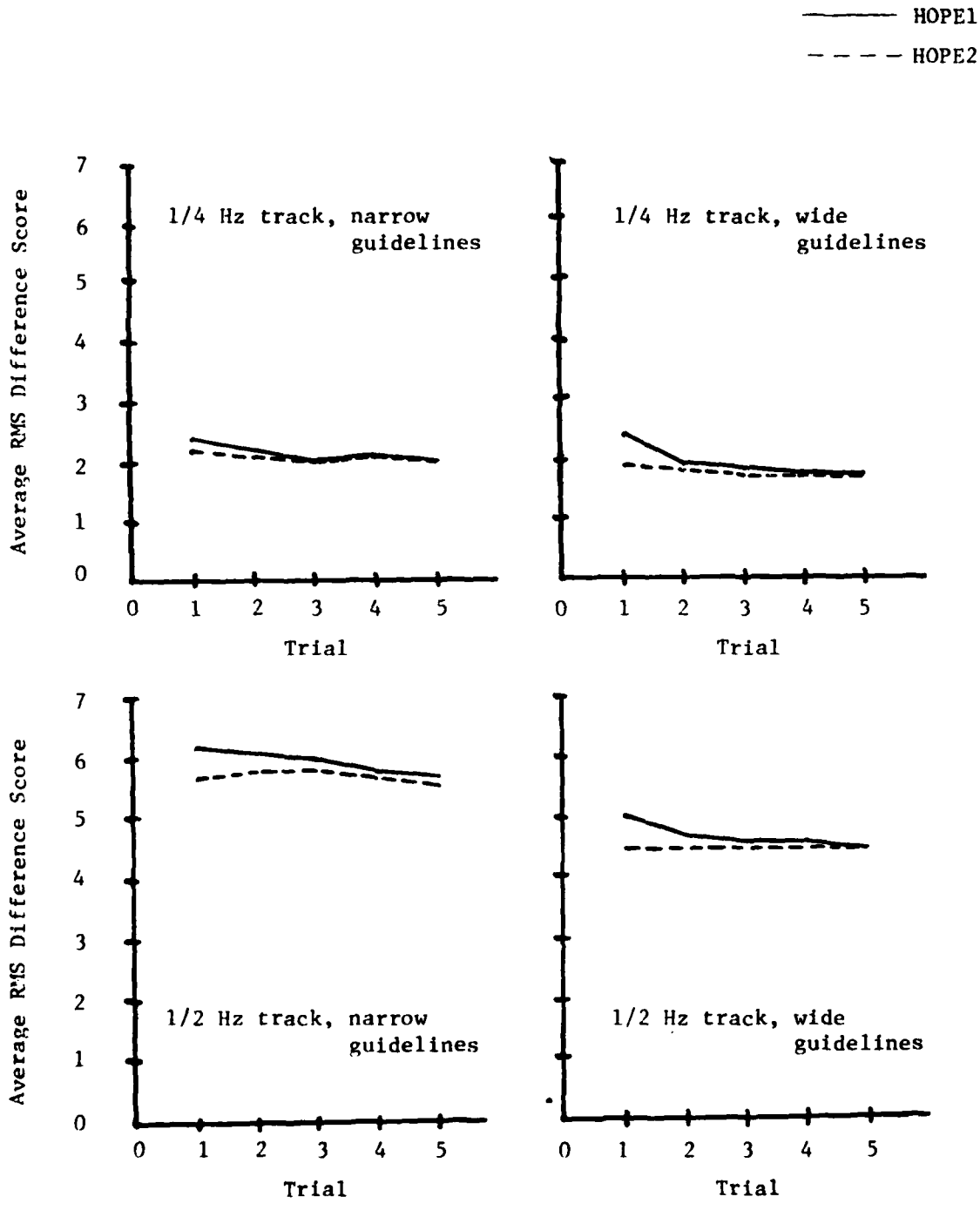


Figure 3. Average Root Mean Square (RMS) Difference Score for Best-Fit Models for Each Trial for Each Condition.

D. Model Refinements

1. Approach to Model Refinements

The approach taken in refining HOPE was to focus initially on reducing the excessive variability in HOPE's control stick positions. This feature certainly made HOPE behavior different from that of humans, and also probably caused problems in the best-fit model identification process, especially given the RMS difference score used to select best-fit models. The RMS difference score reflects only model-human position differences, not velocity or acceleration differences. For this reason, it underestimates the differences between model and human behavior due to model variability and spikiness, which can be thought of as rapid changes in model control stick velocity. In some cases, therefore, the best-fit model dictated by a minimum RMS difference score value was a model which appeared to be considerably more variable than the human behavior it was supposed to match.

Refinements aimed at reducing model variability were, therefore, desirable in two respects. First, reduced model variability would make models look more like human behavior. Second, reduced variability would reduce the effects of shortcomings in the RMS difference score for choosing a best-fit model that looked similar to human behavior.

The approach taken to reduce model variability was to re-examine the basic processes in HOPE to see which of them might be contributing to this problem. Processes associated with the development of the Command Memory and processes which selected commands from that memory were of special interest. Their basic operation and their psychological validity were re-examined to see how they might be modified so that better commands (i.e., commands resulting in less variability) might be developed and selected. One important question was, do these processes operate in a way comparable to the way human processes are believed to operate? If data on human processes were unavailable, the question became, do these processes operate in a logical and efficient fashion?

Based on this investigation, some important changes were made in the operation of the Perception Process, the Excessive Error Process, and the Satisfactory Command Search Process. The changes also affect the Stimulus-Response Association Process and the Command Execution Process. The nature of these changes is detailed below.

2. Modification of the Perception Process

In HOPE1 and HOPE2, there is a distinction between what will be called the "true" cursor position and the "perceived" cursor position. The perceived cursor position is that which the model observes as cursor positions. It is an integer number between 1 and 128. The inability of HOPE to observe cursor positions in a more continuous fashion reflects human thresholds for distinguishing distinct points in visual space. The integer nature and range of perceived cursor positions also correspond with the structure of the Command Memory (a 128 by 128 array), so that cursor positions can be used to address the memory.

The true cursor position is that generated by sending a control stick position through the plant dynamics. It is a real number. The true and perceived cursor positions were related in HOPE1 as follows:

> Perceived cursor position = IFIX (true cursor position) = integer truncation of true cursor position.

This means, for example, that true cursor position between 64.000 and 64.999 were observed by HOPE1 as position 64. In HOPE2, true and perceived cursor positions are related as follows:

> Perceived cursor position = IFIX (true cursor position + .5) = integer rounding of true cursor position.

Thus, true cursors positions between 63.50 and 64.499 are perceived as cursor position 64.

This change was desirable for two reasons. First, it is probably more consistent with human perceptual processes. The human operator is more likely to associate an true cursor position of 64.87 with a perceived cursor of 65 than with 64. Also, rounding instead of truncating results in "more reasonable" commands in the Command Memory. For example, suppose true cursor position is 64.25 (perceived as 64 by the model both before and after the proposed change). Also, suppose that the next desired cursor position is 65. In HOPE1, any command (control stick position) from 72 to 89 resulted in a perceived cursor position of 65 (in one time interval). With rounding, this range is altered to 67 to 76 and 78 to 82 (the jump is due to the way in which nonlinearity in plant dynamics was implemented in this research). This is a result that better reflects the true cursor dynmamics. Thus, through refinement of the Perception Process, more accurate commands are stored in the Command Memory.

3. Refinement of the Satisfactory Command Search Process

In designing HOPE1, one of the most difficult mental processes to represent was that involved in locating or generating a command when a previously unencountered situation arose. There is little available human data or theory on this subject. This process is represented in HOPE by the Satisfactory Command Search (SCS), which in HOPE1 involved a sequence of three operations to locate a command (see Engler et al., 1980). The three operations were as follows:

- a search of Command Memory in the location addressed by the desired state and the last predicted state (assumed to be the current state at command execution time),
- if a command was not found, "column" and "block" searches within the Command Memory were used to try to locate a satisfactory command, and
- the final alternative, if no command had been found, was to use a best-guess command which was obtained by using the desired state as a satisfactory command.

These operations were selected because they seemed to be an efficient way to locate a satisfactory command given the design of the Command Memory. A satisfactory command is one which most closely achieves the desired state. Also, the best-guess command is the ideal command for controller dynamics having no lead or lag characteristics. However, closer examination of the relation in size and consequence of proximal commands in the Command Memory suggested the use of what is believed to be a more psychologically valid procedure for choosing a satisfactory command. This search procedure involves a representation of generalization by the human from knowledge of known commands to inference of a command appropriate to a newly encountered state transition. It involves use of what is called a diagonal search to locate a command which is used to generate (infer) a satisfactory command.

The diagonal search involves searching memory cells on a diagonal which intersects the cell addressed by the last predicted and desired states. The diagonal that is searched is the one whose commands cause state transitions of the same size and direction as that between the last predicted and desired states. The way in which the desired state is defined has been improved in HOPE2 to specify more accurately a true desired state. If the Command Buffer is not empty or will not be empty when the SCS is completed, the desired state is defined as the track point that will be current when the commands already in the buffer and the command to be generated have been executed. If the buffer is empty or will be empty before the SCS has terminated, the desired state is defined as the track point current when the SCS and execution of the command to be generated are complete. These definitions insure the appropriateness of the desired state used in command development. They also represent the human use of preview track information in selecting commands. HOPE2, like humans, recognizes that command development and execution take time, and thus aims at an appropriate future track point rather than at the next track point, which will be outdated before the next command execution is complete.

Figure 4 illustrates the diagonal search that might take place if the desired transition were from cursor position 48 to cursor position 53, and the corresponding memory cell were empty. The letters in the cells indicate the order in which the diagonal would be searched. Each searched cell corresponds to a command causing a state transition similar in size and direction to that desired, although not between the states of interest. The search along the diagonal is terminated when a command is found in a searched cell, or when the search extends beyond a permissible distance along the diagonal away from the cell addressed by the last predicted and desired states. In HOPE2, the diagonal search ends at cells addressed by the desired state ± 20, and the last predicted state ± 20. The limit to the diagonal search represents the fact that human inference is not likely to be based on conditions too dissimilar from the conditions of interest.

The first command found in the diagonal search is used in generating the command that will be put in the Command Buffer. This generation occurs as follows. The difference between the command that is found and the desired state addressing it is calculated. This value is added to the actual desired state and is used as the satisfactory command.

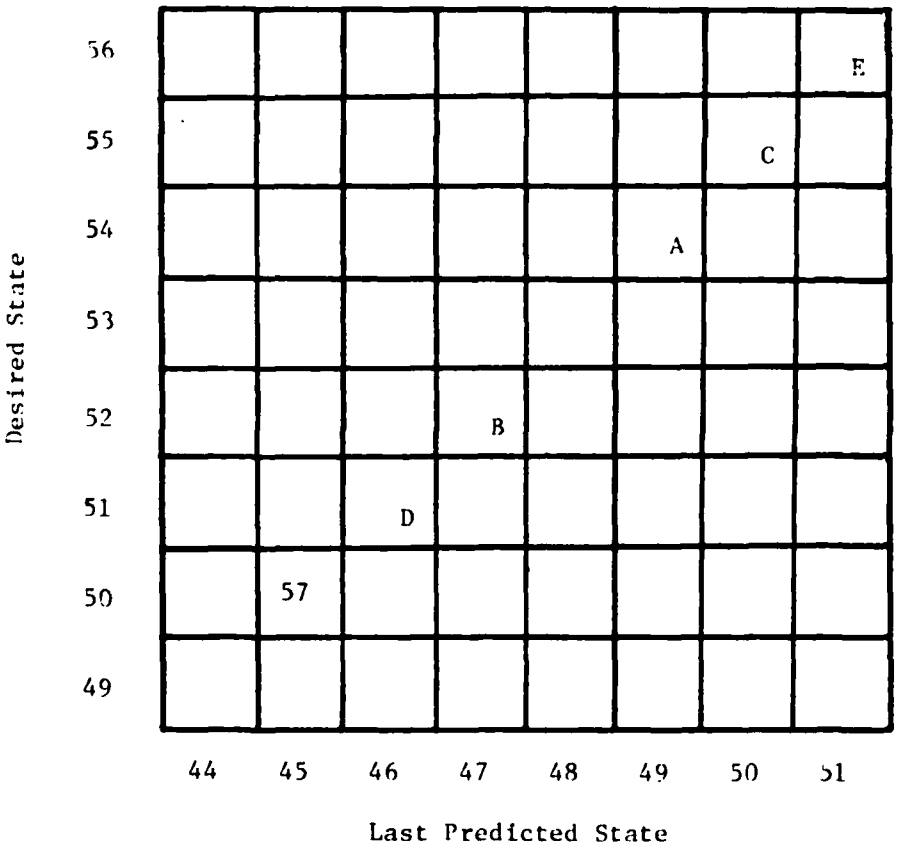


Figure 4. An Illustration of a Diagonal Search--The **desired state is 53 and last** predicted state is 48. The letters in the cells indicate the order of the search. The presence of numbers in the cells indicates stored commands. In this case, the search would terminate with search of the cell addressed by (50,45) since that cell contains a command which can be used to infer the needed command.

This calculation (detailed in Table 1) represents the human ability to use inferences based on known information to deal with newly encountered situations. HOPE2 uses knowledge of commands appropriate for state transitions of a desired size and direction to develop commands appropriate for a specific state transition of that size and direction.

TABLE 1

ILLUSTRATION OF HOW THE DESIRED COMMAND IS INFERRED FROM A COMMAND LOCATED IN THE DIAGONAL SEARCH

Desired State = 53

Last Predicted State = 48

Located Command = 57

Address of Located Command:

Desired State = 50

Last Predicted State = 45

Desired Command = Desired State + (Located Command

- Desired State Address of Located Command)

= 53 + (57 - 50)

Desired Command = 60

This procedure, used in HOPE2 for selecting a satisfactory command, is superior to the column and block searches because it produces a command that may actually accomplish the desired transition and one that is in the correct direction, at least. The procedure for doing this represents human generalization from known information. The column search used in HOPE1 could never move the cursor to the desired state, although the achieved state was predictable. The block search of HOPE1 involved selection of commands corresponding to different states from those of interest, and the result of applying the outcome of a block search was highly unpredictable.

In HOPE1, if neither the block nor column search located a command, the final resort was the use of a best-guess command. The best-guess command in HOPE1 involved use of a "position guess" where the command selected for execution was the desired state (i.e., desired position). This assumes that in the absence of other knowledge concerning an effective command, humans choose to move their control stick to the desired cursor position. It does not, however, reflect the fact that humans probably quickly recognize the lag in the plant dynamics and learn that effective commands must, therefore, have values which lead the desired state. They probably use this knowledge in formulating a best-guess command. In HOPE2 the accuracy of the best-guess command has

been increased by making the best guess equal to the desired state modified by one multiple of ADJUST. ADJUST is added to the desired state if the last predicted state is less than the desired state, and subtracted if the opposite is true. Use of the desired state modified by one multiple of ADJUST represents application of knowledge of plant dynamics to choice of a best-guess command.

To summarize, the Satisfactory Command Search has been modified in HOPE2 to represent human use of inference in generating commands for new conditions. The operations which occur in HOPE2 are listed below:

- entrance to the Command Memory at the location addressed by the desired state and the last predicted state,
- a diagonal search around that location to find a command associated with a state transition of the same direction and size as that desired, and use of this command if located to infer a command effective for moving to the desired state of interest, and
- use of a best-guess command consisting of the desired state modified by one multiple of ADJUST.

4. Refinement of Excessive Error Process

A variety of changes were made to refine the Excessive Error Process (EEP). First, the duration of the EEP was increased. In HOPE1, the EEP was estimated to take 40 msec. Further consideration of the operations involved in the EEP suggested that this estimate was too short. The EEP involves at least a full cycle of perception, response selection and movement initiation. Consideration was given to estimating the EEP as having the duration of a simple discrete choice reaction time, generally estimated as requiring about 200 msec. However, the present task involves continuous movement, probably shortening the response time. It was decided to estimate EEP time as 120 msec, which is also the time estimate used for the Satisfactory Command Search. These processes are similar, especially as represented in HOPE2, and, therefore, similar time estimates should apply.

A second change in the EEP was an increase in the sophistication of its command search procedure. In HOPE1, this procedure involved a search of the appropriate memory location and, if that failed, the use of the same type of best-guess command as that used for the Satisfactory Command Search. In HOPE2, the Excessive Error Process uses procedures similar to those used in HOPE2 for the Satisfactory Command Search. First, there is a search of the Command Memory location corresponding to the desired state and last predicted state. The desired state is that track point which will be appropriate at the end of the time interval necessary for completion of the EEP and execution of the selected command. Use of this track point as the desired state requires HOPE2 to choose a track point occurring some time in the future, and represents the human use of preview information in the current selection of commands. The last predicted state is the predicted outcome of last generated

command. Failure to find a command in the appropriate memory location is followed by the same type of diagonal search as that used in the Satisfactory Command Search. The procedures used there are used in the EEP to infer the command appropriate for the desired state transition.

If the diagonal search does not locate a command, then a best-guess command is used. The procedure used by the EEP to generate a best-guess command is similar to that used by the Satisfactory Command Search, but more complex in that the adjustment of the position guess (the desired state) can vary in degree and is a function of whether the same type of error is being repeated. Each time a best-guess command must be developed, there is analysis of whether the cursor is lagging or leading the track with respect to the desired state. The first time the EEP is called (following a series of states not in excessive error), the best guess command consists of the desired state modified by one multiple of ADJUST. ADJUST is added to the desired state if the current state is less than the desired state (i.e., cursor is lagging). ADJUST is subtracted from the desired state if the current state is greater than the desired state (i.e., cursor is leading). If the outcome of this best-guess command is used again to correct the same type of error as occurred just previously (i.e., lag or lead), the second position guess is modified by twice the value of ADJUST. Whether the adjustment is added or subtracted to the desired state again takes into consideration whether the cursor is lagging or leading the track. If this second best-guess command still leaves the model in excessive error of the same type, a third position guess is modified by three times the value of ADJUST.

A third area of change from HOPE1 is in the events following an adjusted position guess. In HOPE1, a predicted result of a best-guess command was used as the new starting point for the Command Selection Process. Since the result of a best-guess was unpredictable, this introduced the potential for error in subsequent Command Selection. Also, in HOPE1, re-enabling of the EEP was delayed to allow time for perceiving the effect of the command provided by EEP. The delay was implemented by requiring that two Stimulus-Response Associations occur before re-enabling. If the EEP is allowed to attempt to correct for excessive error before the results of a previous attempt are known, the process is not closed-loop and can become unstable. The occurrence of two Stimulus-Response Associations allows enough time for the EEP command to execute.

In HOPE2, the Supervisory Processor is responsible for re-enabling both the EEP and the Command Selection Process following excessive error. Neither of these are re-enabled until the consequence of the best-guess command is available. This allows the Command Selection Process to use the true outcome of the best-guess command as the basis for continued selection of commands. If ineffective commands are still being selected, the EEP can interrupt to initiate corrective action.

A fourth area of change associated with the EEP is with respect to the dumping of the Supervisory Processor request queue. In HOPE1, if the EEP was called, this led to a dumping of the request queue. It was assumed that occurrence of excessive error signified that plans

for other processes should be re-developed. However, the dumping of the request queue included the dumping of requests for Stimulus-Response Association. This caused the loss of information about commands and associated cursor positions occurring just prior to the excessive error. To reduce this loss the Supervisory Processor in HOPE2 processes one requested Stimulus-Response Association before the EEP begins and the queue is dumped.

A final change related to the EEP is that it can now interrupt the Command Execution Process if excessive error is detected. In HOPE1, commands were executed for their full Command Operative Time, even if excessive error was detected during this time. This delayed measures to correct the error. In HOPE2, corrections for conditions of excessive error can begin as soon as a course of action has been decided upon by the EEP.

E. Outcomes of Model Refinements

1. Reduction in Differences Between Model and Human Control Stick Patterns

The major thrust of the model refinements was towards reducing model variability so that model behavior was smoother and more continuous, like that of humans. The effectiveness of the refinements was evaluated by examining whether they reduced model variability, as evidenced by greater smoothness in model behavior and improved matching of human behavior by the model. This latter criterion was especially important, given the assumption that model variability decreased the matching ability of the model. The effects of the attempts at reducing model variability were also evaluated with respect to their ability to relieve the other problematic data patterns described in Section IIC.

A variety of evidence suggests that HOPE2 behavior is less variable and matches human behavior better than did HOPE1. Figure 5a shows a plot of representative behavior for a model generated by HOPE1. The variability and spikiness of behavior are apparent. Figure 5b shows a plot of the same model in the same time bin as generated by HOPE2. The variability and spikiness of behavior are reduced in HOPE2. The reduced variability in model behavior was achieved without increases in errors in model behavior. Table 2 contains some of the diagnostics associated with a representative model as generated by HOPE1 and by HOPE2. Both the position error and the number of requests for the Excessive Error Process are decreased in HOPE2.

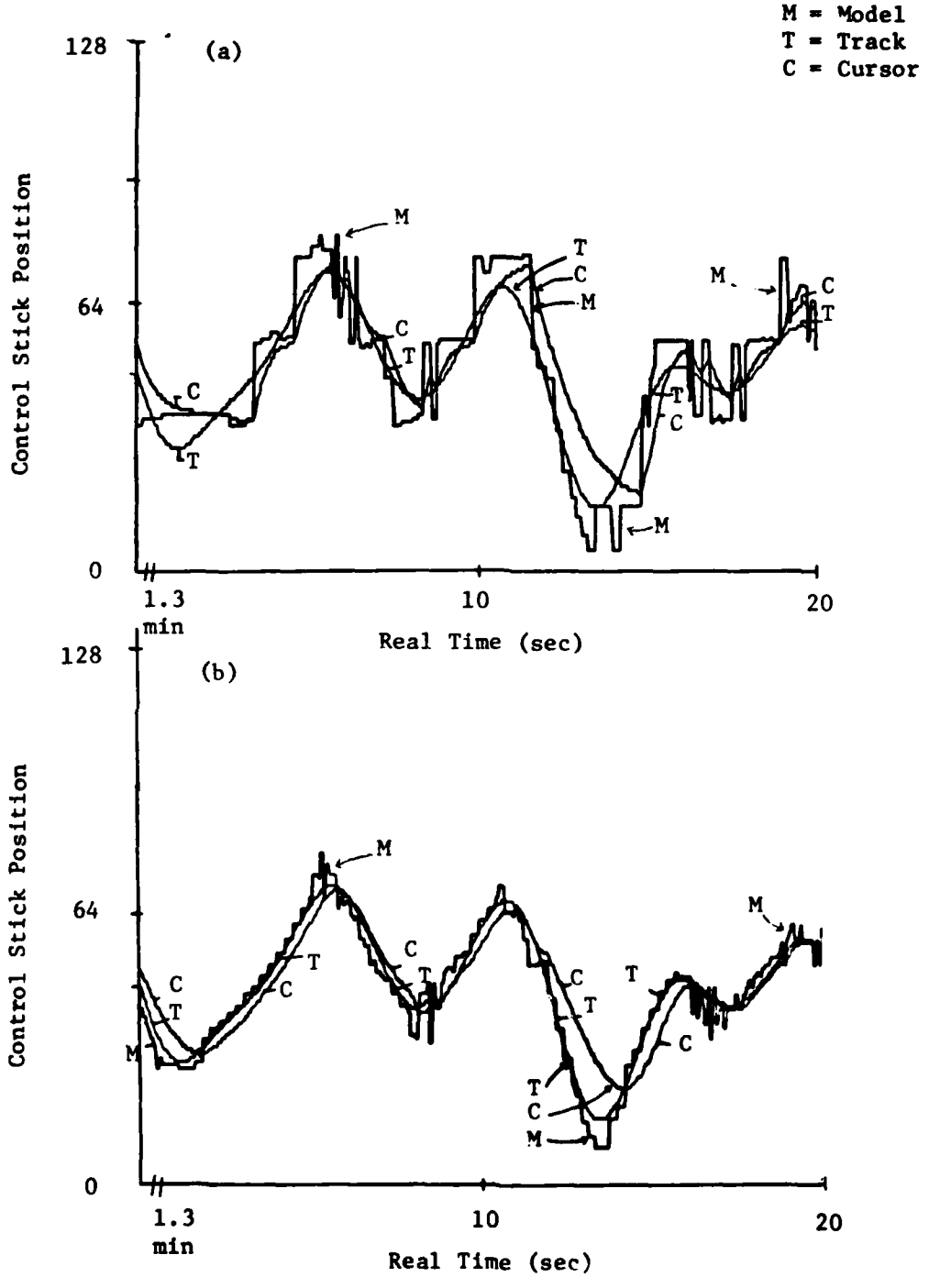


Figure 5. Comparison of a Model as Generated by HOPE1 (a) and HOPE2 (b) for the Same 20 sec Time Bin of Tracking a 1/4 Hz, Wide Guideline Track. HOPE2 Model Behavior is Less Variable and Spiky Than is HOPE1.

TABLE 2

COMPARISON OF ERROR-RELATED DIAGNOSTICS ASSOCIATED WITH A REPRESENTATIVE MODEL (MODEL 41) GENERATED BY HOPE1 AND HOPE2

Time Bin (40 sec each)	Absolute Position Error		Number of Calls to Excessive Error Process	
	HOPE1	HOPE2	HOPE1	HOPE2
1	7.21900	2.67000	18	0
2	5.29200	2.55300	5	0
3	4.90100	3.16300	12	4
4	3.59300	2.65000	2	0
5	3.60200	2.77800	1	0
6	5.78100	4.10000	15	0
7	2.99300	2.45000	0	0
8	2.96800	2.77900	0	0
9	3.82200	3.05700	9	0
10	3.39200	2.72700	4	0
11	3.07800	3.12400	0	0
12	5.05700	4.07700	11	0
13	2.92600	2.65500	0	0
14	2.94400	2.51300	0	0
15	3.68200	3.01500	7	0
16	3.41300	2.93800	2	0
17	3.32400	3.13300	0	0
18	5.43900	4.05100	11	0
19	3.04800	2.80200	0	0
20	2.85100	2.70900	0	0
21	3.74500	3.32500	7	4
22	2.98400	3.08000	0	0
23	3.31400	3.46000	0	0
24	5.28700	4.27800	3	0
25	2.87000	2.94500	0	0
26	2.71400	2.90400	0	0
27	3.44400	3.48500	7	2
28	3.51900	3.01900	0	0
29	3.35300	3.31000	0	0
30	5.28300	4.45200	1	0

2. Improvement in Quality of Matches to Human Behavior

HOPE2 is also able to better match human behavior than was HOPE1. This is evidenced by the reduced RMS difference values for HOPE2 best-fit models compared to HOPE1 best-fit models. Figure 3 shows the average RMS values for best-fit models of HOPE1 and HOPE2 for each trial and for each condition. RMS difference values are generally lower for HOPE2.

3. Increased Similarity in Quality of HOPE Matches to Human Behavior in Different Conditions

Figure 3 suggests that several of the problematic matching symptoms discussed in Section II-C have been relieved by the model refinements. The RMS difference values are especially lowered for early trials, and for ½ Hz track conditions. This suggests that HOPE2 can match more equally changes in human behavior that occur with learning or varying training conditions. However, even with HOPE2, RMS difference values are still lower for behavior in ¼ Hz track conditions, suggesting areas for further improvement in HOPE or in the matching procedure.

F. Summary

This section described, in general terms, the structure and operation of HOPE and then detailed efforts aimed at refining HOPE. The major thrust of refinement was aimed at reducing model and human behavior differences in control stick position variability, and at improving the model's ability to match more equally human behavior at different points in training and in different training conditions. Model refinements were successful at achieving these aims.

SECTION III

REFINEMENTS TO CONTROL STRATEGY MEASUREMENT PROCEDURE

A. Introduction

1. Summary of Basic Approach to Measurement

The HOPE simulation is designed to measure the time-varying control strategy used by humans during continuous control behavior. The current version of HOPE measures control strategy during preview tracking, an example of continuous control behavior. During experimental tests, humans used a control stick to guide a cursor along an externally-viewed preview track. Their behavior in the form of control stick positions is recorded every 40 msec. HOPE operates on a numerical representation of the track, and outputs control stick positions every 40 msec.

The basic approach for using HOPE to measure human control strategy is summarized below:

- HOPE operates on the numerical representation of the track followed by humans. HOPE is run multiple times, each time using a different set of control strategy parameter (CSP) values. Each run generates a set of control stick positions over time representing a HOPE model of human behavior guided by a particular control strategy.

- The human behavior to be measured in a given time interval (also referred to as a time bin) is compared to the behavior of the different HOPE models in that interval. For each HOPE model, a difference score representing its difference from the human behavior is computed.

- The HOPE model with the smallest difference score is designated the best-fit model for the human behavior in that interval.

- The CSP values of the best-fit model are inferred as representing human control strategy in that interval.

This approach to measuring human behavior is a novel one, but it has a variety of advantages over other currently used approaches for measuring continuous control behavior (see Engler et al., 1980, Section II for discussion). The following discussion assumes that the basic approach is useful and valid. The purpose of this section is to describe refinements in the basic approach that were instituted to improve the quality of the obtained measures of human behavior.

2. Variables Manipulated in Refining the Measurement Procedure

The process of refinement focused on several variables critical in applying the measurement procedure. Each variable, its importance and its form in analysis of the preliminary test results are indicated below.

a. Control strategy parameter (CSP) ranges--The CSP ranges are the ranges of the CSP values used in the different runs of HOPE to produce models of strategy-controlled behavior. In analysis of the preliminary tests of HOPE (see Engler et al., 1980, Section V), the CSP ranges used were as follows: Command Operative Time (COT) = 40, 80, 120, 160, or 200 msec; ERRLIM = 2, 4, 8, 16, or 32 screen units (1 screen unit = .29 cm); ADJUST = 2, 5, or 8 control units (1 control unit = .8% of the total range of control). This means that there were 75 (5 x 5 x 3) models of behavior generated by HOPE and used to measure human control behavior. The CSP ranges and values determine the number of models and the range of model behaviors that HOPE is able to generate and, therefore, the range of strategy-modulated human behavior HOPE is able to measure. Ideally, the CSP range should include all values used by humans in the conditions of training and practice in which they are being measured.

b. Time interval for measurement--Human control strategy is believed to vary over time with learning, so measurement is performed for behavior within specified intervals of time. For preliminary test analysis, behavior was measured for each 20 sec time interval. The interval selected determines HOPE's ability to measure the time-varying nature of control strategy. Ideally, the time interval for measurement should correspond to average duration of the interval over which control strategy is constant in humans.

c. Difference score for gauging model-human matches--The difference score should validly reflect the difference between model and human behavior in a given time interval. Behavior from both model and human can be characterized as discretized waveforms. Ideally, the difference score between the two waveforms should include position, velocity and acceleration differences. However, there does not currently exist a difference score measure that includes all of these aspects of difference. In preliminary tests, the commonly used root mean square (RMS) position difference score was used. This score is the square root of the sum of the squared position differences between model-human behavior within the time interval of measurement.

d. Matching criterion for inferring human control strategy--This criterion dictates the conditions under which best-fit model CSP values will be used to estimate human control strategy. For the preliminary tests, no criterion was instituted. The CSPs of each best-fit model were used to estimate human control strategy. Ideally, the criterion should allow such estimates to be made only when model behaviors are adequately similar to human behavior to capture the human control strategy.

B. Refinements in the Measurement Procedure

The analysis of the preliminary test data was the first application of HOPE to measurement of human control strategy. Before this analysis, there was little basis for making decisions about CSP ranges, time interval for measurement, etc., because the approach was so novel. However, the outcomes of the preliminary test analysis, as well as some outcomes of model refinements (see Section II), suggested some ways in which the measurement procedure might be improved. Areas of change are described in the following discussion.

1. CSP Ranges

Several items suggested need for changes in the CSP ranges. First of all, the COTs of the best-fit models for human behavior in certain conditions of preliminary testing were near the upper limit of the COT range used. That is, many of the best-fit models had COTs of 160 or 200 msec. Such clustering of CSP values near one of the limits (upper or lower) of the range used suggests the need for expanding the CSP range beyond that limit. In the case of COT, the clustering of best-fit model COTs near the upper limit suggested that HOPE models might include a longer COT. Therefore, in validation testing, COT range was expanded to include COTs of 40, 80, 120, 160, 200, and 240 msec. The shorter COT values were retained because they modulated the best-fit models for certain other conditions of preliminary testing.

The ERRLIM range was changed because of the observation that human error rarely reached the magnitude implied by the higher values of the ERRLIM range used in preliminary test analysis. For example, an ERRLIM of 16 screen units (4.64 cm) implies a position error much larger than that observed in human behavior. It was felt that a greater proportion of human behavior could be better modeled by expanding the number of small ERRLIM values and excluding the highest values of ERRLIM. For this reason, ERRLIMs of 0, 2, 4, 6, 8, and 10 screen units were used in data analysis for validation testing. These correspond to values of 0, .6, 1.2, 1.8, 2.4, and 3.9 cm.

The ADJUST range and values within that range were modified to increase the precision with which human ADJUST might be measured. ADJUST values used for validation testing were 2, 4, 6, 8, 10, 12 control units. This allowed precise measurement of small ADJUST values (e.g., 2, 4), and also slightly expanded the upper limit of the range. The latter change was justified by the clustering of ADJUST values for best-fit models at the upper range limits for certain conditions of preliminary testing.

2. Time Interval for Measurement

For preliminary test analysis, human behavior was measured for each 20 sec interval. Both observation of human behavior patterns and discussion with human subjects suggested that control strategy changes were likely to be occurring more often than every 20 sec. It was theorized that human control strategy probably changed no more frequently than did track frequency. One cycle of the faster, ½ Hz cut-off frequency track takes, on the average, 2 sec. The initial

strategy for determining a better matching interval was to examine the possibility that human control strategy was varying near the maximum theorized to be possible, every 2 sec. If model-human matches improved with this shorter interval, this would suggest that the 2 sec interval more validly reflected the interval over which human control strategy was constant.

RMS differences scores for best-fit models using the refined HOPE, 20 sec interval were compared to those using the refined HOPE, 2 sec interval. The average RMS difference values were reduced by nearly half when the 2 sec interval was used. Analyses using intervals intermediate between 20 sec and 2 sec (e.g., 5 sec) did not result in as much improvement in the quality of matching as that observed with a 2 sec interval. It should be noted that although use of the 2 sec interval reduced the average RMS difference scores, it did not uniformly improve matching for all behaviors to be matched. There were instances where the RMS differences scores for the best-fit model for 2 sec intervals were greater than any of those computed with the 20 sec interval. This point will be relevant to upcoming discussion of the criterion for inferring human control strategy.

3. Difference Score for Gauging Model-Human Matches

As mentioned previously, the RMS difference score commonly used takes into account only position differences in describing model-human differences. Comparison of model and human behavior from preliminary test analysis indicated that one of the more striking differences between the behavior of some models and humans was a difference in the velocity of control stick behavior. Humans tended to use relatively smooth, continuous, behavior. They tended to use low velocity movements; they did not make drastic shifts in the size or direction of their movements. Many models showed high velocity movements resulting in the variability and spikiness in behavior discussed earlier (see Section II). It seems important to choose best-fit models which resemble human behavior both in position and in at least velocity, if not also acceleration of movement. Unfortunately, there does not currently exist a difference score which includes position, velocity, and acceleration difference components. In refining the measurement procedures, some effort was devoted to developing such a measure. One candidate is a measure named the Mean Absolute State Error (MASE) which sums position, velocity, and acceleration differences in one measure. Several variations of MASE were given preliminary testing, but could not be adequately developed and tested within the time and budgetary constraints of the research.

At least two problems require careful attention if a valid version of MASE is to be developed. First, it will be necessary to develop equivalent scales for position, velocity and acceleration measures so that they can be meaningfully summed into one score. Secondly, it will be necessary to determine whether each component is to be equally weighted in such a summed score. Completed efforts suggest that best-fit models may be best selected with unequal weightings, giving position difference the greatest weight, and velocity and acceleration differences somewhat less weight in determining the overall difference score.

However, a satisfactory scheme for determining appropriate weights still remains to be devised.

It should be emphasized that the characteristics and quality of the difference score used to select best-fit models is of critical importance to the quality of the measures of human behavior obtained using the present research approach. Developing a new measure is a major research effort in itself which could not be accomplished within the present program. Further, results from analyses using the RMS difference score compared to several preliminary versions of MASE revealed no striking differences in the patterns of CSP estimates. Preliminary versions of MASE were too crude to improve upon control strategy measures obtained through use of the RMS difference score. Therefore, it was decided that the refined measurement procedure would continue to use the RMS difference score to select best-fit models. This measure is commonly used to assess waveform differences, and is the best measure available for present purposes.

4. Matching Criterion for Inferring Human Control Strategy

In preliminary testing, the CSPs of all best-fit models were used to infer human control strategy. However, careful examination of the similarity between best-fit models and human behavior in certain intervals indicated that in some cases even the best-fit model was a poor approximation for the human behavior to be measured. This became especially apparent once 2 sec matching intervals were used. Although on the average, use of the 2 sec interval improved the overall quality of model matching, there were some intervals of human behavior that were poorly matched, as indicated by the large RMS difference scores of even the best-fit models.

Based on these observations, it was decided that human control strategy would not be inferred from the best-fit model unless the quality of matching was very high. Visual examination of plots of model and human behavior suggested that model behavior was acceptably similar to human behavior if the RMS difference score was 12.8 or less. (The score is in terms of units of control, where there are 128 possible control positions.) Therefore, in analysis of validation test data, it was decided that the RMS difference score has to be 12.8 or less if the CSPs of the best-fit model are to be used to infer human control strategy. Best-fit models that have larger scores are not adequately similar to human behavior for human control strategy to be inferred. This requirement is the "quality" component of the criterion for inference-making.

Analysis of preliminary test data also revealed that for a small proportion of the intervals to be measured, the RMS difference scores were equal for the first and second best-fit models. This made it unclear which model CSPs should be used to infer human control strategy, and pointed out an additional problem with use of the RMS difference score. To deal with this difficulty, it was decided to add a second component to the criterion for inferring human control strategy. The criterion now requires that: a) the best-fit model must match human behavior with an RMS difference score of 12.8 control units or less,

and b) the best-fit model RMS difference score must be lower than that of the second best-fit model, as distinguished by a difference in at least the first decimal (tenths) place of the RMS difference score. This latter component of the criterion is referred to as the "uniqueness" component.

C. Summary of Refined Measurement Procedure

For the validation test analysis, the measurement procedure was refined to improve the quality of model matching to human behavior, and the quality of inferences about human control strategy. The CSP ranges were revised to allow HOPE to produce models more representative of the range of behaviors exhibited by humans in the experimental conditions of testing and practice. The revised CSP ranges are as follows: COT = 40, 80, 120, 160, 200, or 240 msec; ERRLIM = 0, 2, 4, 6, 8, or 10 screen units; ADJUST = 2, 4, 6, 8, 10, or 12 control units. It was decided to match 2 sec intervals of human behavior since that interval seemed more representative of the duration over which humans held control strategy constant, and greatly improved the average quality of the model matching. It was decided to continue use of the RMS difference score for selecting the best-fit model, although future efforts should be devoted to developing a measure more descriptive of model-human differences. Finally, a criterion for inferring human control strategy from best-fit models was imposed. Human control strategy will be inferred from the CSPs of a best-fit model only if the model has an RMS difference score of 12.8 or less, and if its score is lower than that of the second best-fit model.

SECTION IV

VALIDATION TESTING AND RESULTS

A. Introduction

The concept of control strategy is the central theme in this research. Control strategy is the object of measurement by the simulation HOPE. Control strategy is believed to determine or define the functioning of the cognitive processes believed important in learning continuous manual control tasks. Through that determination control strategy profoundly affects the style and quality of human performance.

A parameter is a variable whose value determines the characteristics or behavior of a process. Control strategy is defined as the set of parameter values that determine the functioning of the processes important in continuous manual control. There are three categories of control strategy parameters. These three categories are:

- criteria for performance in all aspects of each subtask of the overall task,
- stimulus cues on which to base performance, and
- sequence for decision-making processes.

Each of these types of parameters influences the mental processes important in continuous manual control learning and performance in distinct ways. Criteria for performance provide a basis for a variety of comparisons important to motor skill learning. These criteria dictate, for example, standards for acceptable operator behaviors (e.g., timing, boldness) as well as standards for the controlled system's outputs (e.g., allowable error). Selection of stimulus cues determines which information the operator will perceive and remember for use as a basis for motor performance. The sequence for decision-making processes determines the order in which processes such as developing responses to excessive error may occur or to novel situations. The theory underlying this construct is presented in more detail in Section III of Engler et al. (1980).

1. Purposes of Testing

The testing described in this section was carried out for the following purposes:

- to further validate the procedure used to identify human control strategy,
- to further validate HOPE's ability to measure changes in control strategy that occur over the course of task learning, or that occur between training conditions, and

- to validate the predictive validity of measures of control strategy made using HOPE.

The validity of a measure of a psychological construct is the appropriateness of the inferences made from such measurement (Standards, 1974). Inferences made from measurements taken in this research are of two general types. The first type of inference is related to the object of measurement--the construct. In the present context, investigations of construct validity involve examining whether the obtained measures actually measure human control strategy. This is, validation involves efforts to discover how faithfully the test measures represent the construct being measured. The second type--criterion-related validity--involves determining the relationship of the measures taken to other measures. One example of criterion-related validity is predictive validity. The extent of predictive validity is the extent to which an individual's future level on some criterion behavior can be predicted from a knowledge of a prior measurement value. Predictive validity involves a time interval. It may involve predictions of other behavior than that originally measured, or it may involve predictions of later performance on the same constructs measured previously. High school grades, for example, may be accepted valid predictors for freshman college grades.

It is important to remember that validity cannot be directly measured, but is rather inferred from the collections of measurements made in some procedure for validation. Validity may be judged, for example, to be adequate, marginal, or unsatisfactory (Standards, 1974). Validation necessarily involves a variety of investigative processes. The validation tests reported here have used a variety of types of analysis directed toward assessing both types of inference that can be made from these measures of control strategy--the faithfulness of representation and the extent of predictive power.

During the early part of this research, a set of preliminary tests was carried out (detailed in Engler et al., 1980) in order to make a preliminary assessment about the extent to which the construct control strategy is faithfully represented in the measures taken. Since those tests, refinements have been made in the HOPE simulation and in the procedures for utilizing it to measure human control strategy (see Sections II and III). The validation tests described here involve not only examination of the degree to which measures of control strategy made using HOPE are adequate representations of human control strategy, but also examination of the predictive validity of control strategy measurements. The purposes of the validation testing were accomplished by structuring an experiment around four basic research questions. Details of experimental procedure will be described later, but are summarized here. In the validation tests, subjects tracked (with preview) for three trial sets. The first and last sets of five trials were of the same type, either a ½ Hz cutoff frequency random track, or a ¼ Hz cutoff frequency track. The second trial set was the alternate. The tracks and experimental setup are otherwise nearly identical to that used in preliminary testing (see Engler et al., 1980). The research questions are listed and discussed below.

Question 1: Do HOPE models match human behavior well enough to permit estimation of human control strategy?

The procedure for measuring human control strategy in a given time interval involves identifying as the human control strategy that strategy which is used by the HOPE model which best matches the human behavior in that interval. Validation of this measurement procedure requires minimally that HOPE produce predictions of control stick positions that match human behavior to an acceptable extent. The criterion for acceptable matching in the validation tests is more strict than that which was adopted for the preliminary testing, and is as follows. For at least 90% of the subjects, one or more HOPE models must match human behavior with a root mean square (RMS) difference score of less than 12.8 (control position units) for at least 80 percent of the duration of the testing. This difference score requires that HOPE match human behavior within 10 percent of the control stick's range of motion, since there 128 possible control positions. In addition to this quality component of the matching criterion, there must be one HOPE model which is uniquely best (at tenths place precision) in order for any inference about the human control strategy to be made.

In preliminary testing, HOPE matched human behavior acceptably within a more lax quality criterion (RMS differences within 20 percent of the control stick's range of motion; for 50% of the duration of testing). No uniqueness component was utilized in that testing. However, the quality of the matches varied as a function of the extent of human learning and the frequency of the track being followed. HOPE matches were better for later trials of human learning, and for ¼ Hz tracks than for ½ Hz tracks. Refinements of HOPE have been aimed towards reducing these differences. The validation test data will help indicate whether the refinements have been successful. Ideally, HOPE should match human behavior within the two criteria indicated above, and should match equally well for all training conditions and degrees of learning.

Question 2: Does control strategy, as identified by HOPE, change with learning?

Control strategy is believed to change during the learning of a new psychomotor task (see Engler et al., 1980). Most individuals begin a new psychomotor task using relatively ineffective strategies for performance. These strategies may be ineffective because they are poorly defined, or they may be well defined, but based on experiences with other, different tasks. With practice, the initial control strategy is revised, and a strategy tailored for the current task (i.e., a task-specific strategy) is developed.

If this conceptualization of control strategy is correct, then measures of human control strategy should change with learning. The control strategy parameters (CSPs) inferred for humans at the beginning of training should differ from those inferred at the end of training.

For the task utilized in the validation testing, the following developmental trends were predicted. First, it was expected that COT should become smaller with experience, at least when subjects tracked

in the ½ Hz track condition. COT is inversely related to frequency of control movements, at least to the upper limit for frequency of movements. That is, lower values of COT permit higher frequency movements. Frequency of control movements should be related to frequency of presented track. Therefore, COT should also relate to the track conditions presented. Preliminary testing suggests that subjects begin tracking with COT at a level appropriate for ¼ Hz tracking. For these reasons it was predicted that COT would decrease, particularly for ½ Hz track conditions.

The second development expected was that ADJUST values would show changes with learning, rather than being immediately at a task-specific level. Just as in the preliminary tests, the control dynamics involved a lag betwen control input and system output. The lag was greatest at the extremes of the control stick's range of motion and reduced, in a step-wise fashion, toward the center. One of the fundamental assumptions made in this testing was that subjects were naive with respect to these dynamics, although obviously not with respect to tracking tasks in general. The pattern of the variable lag, and therefore full knowledge of plant dynamics, was unknown to them. Positive values of ADJUST reflect the adaptation of subjects to lag--i.e., the slower the controlled-element response, the larger ADJUST should be to compensate for the slowness. (Negative values might be appropriate for controls involving high gain.) This compensation for lag occurs not only in response to excessive error, but also under any conditions in which knowledge of the exact move required is lacking. Subjects in ½ Hz conditions must use the extremes of control stick range (the slower response areas) more often than ¼ Hz subjects. Because the precise control required is initially unknown to them, and because subjects in ½ Hz conditions experience a greater proportion of slow responses, it was predicted that subjects in ½ Hz track conditions would begin with a value of ADJUST like that for those in ¼ Hz conditions, but that as a result of developing experience, one or both groups of subjects would change in such a way as to result in ½ Hz condition subjects using a larger ADJUST by the end of the 15 trials.

The preceding arguments might seem to imply that the direction of change for ADJUST should be positive--that is, ADJUST should increase over time. It should be recalled, however, that the direction of change to some final state such that ½ Hz condition subjects are using larger ADJUSTs than ¼ Hz condition subjects depends upon the initial values estimated for ADJUST. If subjects assume no lag in control, then an increase would be predicted for both groups. If subjects overestimate the compensation needed for lag, or tend to over-react to excessive error early in tracking, then both groups might decrease. The preliminary testing involved the same control dynamics as do these tests, but a different range for ADJUST was used, as well as a slightly different track pattern. Thus, no good estimate of starting values was available.

For similar reasons, a prediction of change in ERRLIM was made, without specification of the direction of the change over time. The absence of a directional prediction is due to the uncertainty about the starting values for ERRLIM that would be used by subjects. This uncertainty results from differences in the range of ERRLIM used in

preliminary and validation tests; and a resulting absence of starting value predictions. It was expected that by the end of training, ERRLIM in the easier ¼ Hz condition would be smaller than ERRLIM in the ½ Hz condition. Values of ERRLIM were expected to be the same in both groups at the beginning and to change in such a way as to result in the easier task group having the smallest ERRLIM.

The easier task was expected to be associated with a smaller ERRLIM because a similar relationship had been found in preliminary testing. Further, such a relationship seems appropriate to the meaning of this parameter. ERRLIM is believed to represent an internal standard by which system performance is evaluated. The better system performance is perceived to be the stricter the standard for evaluation can be without stimulation of an excessive amount of effort devoted to nulling error.

In summary, the following predictions were made about changes in control strategy with learning.

- COT estimates should decrease with experience, particularly for persons in ½ Hz track conditions.

- ADJUST estimates should change with experience, in ½ and/or ¼ Hz conditions, so that by the end of testing, ADJUST values estimated for subjects in ½ Hz conditions would be larger than those estimated for subjects in ¼ Hz conditions.

- ERRLIM estimates should change with experience, in either ¼ or ½ Hz conditions, so that ERRLIM estimates for persons in ¼ Hz conditions are smaller than ERRLIM estimates for those in ½ Hz conditions.

Question 3: Does control strategy, as identified by HOPE, reflect differences between training conditions?

An important assumption underlying the definition of control strategy is that it reflects variations in the training environment. When factors such as task difficulty or available cues change, control strategy gradually changes so that overt behavior can remain fairly effective. For example, a driver can stay on the road fairly successfully in dry weather, or in a snowstorm, if he adaptively modulates his strategy for driving, e.g., his acceptable speed, his accelerations, the environmental factors he attends to, etc.

If the conceptualization of control strategy in HOPE is correct, then control strategy as identified by HOPE should reflect differences between training environments. In the experiment to be described, the training environment was varied to create two distinct training environments, each maintaining the same non-linear, variable lag, control dynamics. Subjects tracked, for 10 of 15 trials, either a more rapidly varying ½ Hz track, or a less rapidly varying ¼ Hz track. These two tracks make different demands on the human operator, just as driving on a curving road makes demands different from driving on a straight

road. For example, motor commands must vary more rapidly when tracking a more rapidly changing track or road. In HOPE, such changes might be reflected in changes in Command Operative Time (COT), which controls the frequency with which motor commands can vary. It was predicted that estimated COTs for subjects in ½ Hz tracking would be shorter than COTs estimated for subjects in ¼ Hz conditions, allowing more frequent execution of new commands. The results of the preliminary testing of HOPE were consistent with this prediction, which received further testing in the present experiment.

A more rapidly curving track, combined with control of an unfamiliar lag-type plant, should also cause subjects to make more energetic responses to excessive error in situations where knowledge of the precise move needed is lacking, since the track is likely to be moving away from the controlled element at a more rapid rate in ½ Hz conditions. This idea suggests that ADJUST values estimated for subjects in ½ Hz track conditions should be larger than those in ¼ Hz track conditions. This prediction also received support in the preliminary testing, and was examined in the second experiment.

ERRLIM values are believed to be related to the human operator's internal standard for performance. Larger values for estimated ERRLIM should result from less strict internal standards. The reader will recall that, in HOPE, ERRLIM determines the maximum distance from the center of the track the cursor is permitted to move before the Excessive Error Process is activated. In preliminary tests, it was clear that the more rapidly varying ½ Hz track was associated with higher human (and model) error--a larger average distance from the center of the track. It seems plausible that in more difficult tracking conditions, subjects might apply a more lenient standard of performance in order to avoid overly-frequent judgements of excessive error, and to avoid making impossible demands on themselves. Thus, subjects in ½ Hz tracking conditions should, on the average, have a larger estimated ERRLIM than subjects in the ¼ Hz conditions.

It is important to note that differences in control strategy may not be immediately apparent, but may emerge during the course of learning. As was discussed with reference to Question 2, initial control strategies, although often related to past experience, are likely to be relatively ineffective in performance of a new task. In the present experiments using subjects inexperienced in the control dynamics of the system, it seemed unlikely that there would initially be systematic differences in the control strategies used by subjects in the two different training conditions. However, if the conceptualization of control strategy is correct, differences in the control strategies used in the different training conditions should emerge over the course of learning, with more clear cut differences emerging on the later trials. For these reasons, special attention was focused on the CSP values estimated from behavior in the first and last trial of the two different training conditions.

In summary, the following predictions were made about CSP differences between training conditions.

- By the end of testing, COT estimates should be larger for subjects in ¼ Hz conditions than for those in ½ Hz conditions.

- By the end of testing, ERRLIM estimates should be smaller in ¼ Hz than in ½ Hz conditions.

- By the end of training, ADJUST estimates should be smaller for subjects in ¼ Hz conditions than in ½ Hz conditions.

- At the beginning of testing, there should be no differences between estimates of control strategy parameters in the two conditions.

Question 4: Do measures of control strategy show predictive validity?

This question is of interest because the long-term goals for this research include the possibility of utilizing control strategy measures as predictors. Control strategy might be measured in a simulator for the purpose of predicting later control strategy or performance, either in a simulator or in an aircraft. Predicting later control strategy would be quite desirable if an optimal control strategy for a certain flight maneuver could be determined. If this were possible, then the ability to predict whether that optimal CS would be used in actual flight, based on measures taken at some prior time in a simulator, would be a valuable aid in determining whether trainees were ready to transfer to the actual aircraft. In general, there is a need to better predict performance in aircraft from measures taken in a simulator. Measures of control strategy combined with performance ratings taken in a simulator might together provide a much more valid predictor of later performance than do performance ratings alone.

These long-term considerations helped focus these preliminary investigations on two types of predictions. The first is the prediction of later control strategy parameters from earlier measures of control strategy. The data from the validation tests allow two examples of prediction of this type. The first involves the use of measures of control strategy taken early during performance in a training condition--such as the ¼ Hz or ½ Hz track conditions--to predict the control strategy used later in the same training condition. One possible predictor measure would be that made at the beginning of training, before a control strategy specific to the training condition has developed. The later tailored strategy may bear some discernible relationship to the earlier measures. In these tests, for example, the control strategy achieved in the last trial of ¼ Hz trials may bear an orderly relationship to that used at initial exposure to the ¼ Hz training condition. This relationship would not, however, be predicted if naive subjects begin training with a random variety of strategies which then merge to a relatively common, task-specific strategy. There simply is not sufficient information on the control strategies used by persons encountering a new task to indicate whether first-used control strategy is a good predictor of last-trial control strategy, but the relationship was investigated in these tests.

Another possible predictor for last trial control strategy is the set of measures made after some experience in the task, but at some period earlier than the end of testing. In the tests here, which involved two separate sets of ¼ Hz trials for one subject group, it would be expected that control strategy measures taken at the end of the first ¼ Hz training condition should be predictive of control strategy measures at the end of the second ¼ Hz training condition. Measures taken after four trials in the ¼ Hz condition should be fairly task-specific. Preliminary tests indicated that control strategy was task-specific after 15 minutes of tracking; four trials in these tests involved 12 minutes of tracking. Therefore, measures taken in the fifth trial were thought to be close to task-specific. These task-specific measures should correlate positively with the task-specific measures taken at the end of training.

A second example of prediction possible between two measures of control strategy is that between measures taken in <u>different</u> tasks or training conditions. In these validation tests, for example, measures taken at the end of a ¼ Hz track condition might be predictive of measures taken at the first, or even the last part of a subsequent ¼ Hz track condition. If such predictive relationships exist at all, it seems likely that the sign of linear correlations should be positive, for the following reasons. Fifth trial control strategy should be task-specific and should be associated with improved task performance, compared to the beginning of testing. (Both these ideas were supported by preliminary testing.) If the assumption is made that improved performance is reinforcing, then the sign of any predictive relationships between the control strategy used successfully in one track condition and the control strategy used in a second, somewhat similar tracking condition, should be positive, since people tend to repeat responses for which they have been rewarded. In the absence of data demonstrating the contrary, linear relationships between control strategy measures made prior and after transfer were expected to be positive.

A second type of prediction suggested by the long-term goals of this research is the prediction of later performance from earlier measures of control strategy. RMS error computed from comparisons between cursor position and track points is a performance measure for the tracking task in these validation tests. If control strategy affects performance, then measures of a stabilized control strategy should be predictive of later performance. In these tests, for example, by the fourth trial of either the ½ or ¼ Hz training conditions, control strategy should be close to a strategy tailored for that training condition. If so, then RMS error in the fourth and fifth trials should show orderly relationships to one or more of those control strategy measures.

Specifically, in ½ Hz conditions, COT should be related positively to error. This is because the more frequent moves associated with low values of COT seem to offer the controller (whether human or HOPE model) the best chance to keep up with the rather quickly moving ½ Hz track. ERRLIM, too, was expected, in ½ Hz conditions, to be positively associated with error. This is due to the definition of ERRLIM as representative of standard by which performance is evaluated. Persons using higher standards (smaller ERRLIMs) should have lower error.

Because the control dynamics used in testing involve a variable lag, compensation for lag (ADJUST) should be related negatively to error, especially for ½ Hz tracking. This condition involves a greater exposure to very slowly responding areas of control, so use of larger ADJUST values should permit lower error.

Predictions about the relationships expected in ¼ Hz conditions are similar, though not identical. COT, for example, may not relate to error in the COT ranges used in estimation. Both small (40 msec) and large (240 msec) values for COT seem quite adequate to keep up with a track that takes, on the average, 4 seconds to go through a complete right to left and return cycle. ERRLIM, however, should be positively related to error, due to its definition as an internal performance standard, even in this simple ¼ Hz tracking condition. ADJUST, because of its function as a compensation for lag, may relate negatively to error in ¼ Hz conditions. The factor which makes these predictions seem less likely than those for ½ Hz conditions is the very narrow range of error recorded for the easy ¼ Hz conditions.

It may also be possible to predict error after transfer from one condition to another from measures of CSPs made prior to transfer. The same directional relationships specified for same condition prediction should hold here, as well--e.g., in ½ Hz conditions, transfer predictions should be positive for COT and ERRLIM; negative for ADJUST. The reason for expecting similar relationships is two-fold. First, the basic similarities of the two tracking conditions result in predictions of similar within-condition control strategy-error relationships. Second, if subjects structure their control strategy based on these relationships, then similar control strategies could be expected to carry over successfully into a new, but similar, tracking condition, as was argued previously. Control strategy prior to transfer may be predictive of control strategy, and therefore error after transfer to a similar condition.

There is still another possible use for a simulation of this nature which serves to suggest tests to be made of the predictive validity of this preliminary version. That is, suppose measures of control strategy taken early in training were used to specify a model of an individual. That model, then, might be subjected to a variety of training conditions in order to prescribe the best condition for the individual who had been measured. In such a situation, the model selected becomes a substitute for the human whose learning and performance are being predicted. An analog to this situation in the current research is that CSP sets selected on the basis of measures taken early in the validation tests might, when utilized in HOPE, produce predictions of human control stick and cursor behavior which would be in some sense acceptably close to those actually produced by the human.

The models selected to represent subjects in the validation tests could be selected on the basis of mean or modal values, based on all CSP measures taken in some trial. These representative models, however, could not represent the developing control strategies which are the focus of interest in this research. Such a representation requires integration of HOPE with a model of the development of control strategy

over time--one of the goals of further research based on the work accomplished here. These fixed-strategy representative models do, however, permit a very preliminary examination of the potential of such a substitution approach.

In summary, the following results may be expected from analysis of the predictive validity of control strategy measures.

- There may be linear relationships between first and last trial control strategy measures.
- There were expected to be positive linear relationships between fifth and last trial control strategy measures.
- There were expected to be positive linear relationships between fifth trial control strategy measures and those taken after transfer to another similar tracking condition.
- COT and ERRLIM should be positively related to error in ½ Hz conditions; ADJUST negatively. ERRLIM and ADJUST should be related positively and negatively, respectively, to error in ¼ Hz conditions.
- COT and ERRLIM, measured in ½ Hz conditions, should be positively related to error in a subsequent ¼ Hz condition. ADJUST should be negatively related. ERRLIM and COT measured in ¼ Hz conditions should be positively related to error in a following ½ Hz condition. ADJUST should be related negatively.
- Representative models, selected on the basis of early CSP measures, should provide acceptable predictions of later control inputs.

B. Method

1. Pilot Testing of Validation Test Design

In response to the research questions described above, a transfer task paradigm was proposed for the validation test and given pilot testing. Pilot tests were conducted to gain further information on two issues.

- How much training is necessary for control strategy to stabilize?
- How does the duration of a break between training sessions affect the nature of transfer?

These issues were of concern because they affected the power of the proposed validation test design to address the research questions of interest. Differences in control strategy between conditions, changes in control strategy with learning, and transfer between conditions might not occur if control strategy did not stabilize within the proposed

20-minute training sessions. Even if stability were achieved, the overnight break between training sessions might cause forgetting of control strategy that would preclude any transfer between conditions.

These concerns could not be addressed by examination of data from preliminary tests because no extended breaks during testing had been given subjects nor were any subjects given more than 20 minutes of training. The issue of stability of control strategy had not been addressed at all. Therefore, pilot tests were designed to address these concerns as well as to test out analysis programs to be used in the transfer task paradigm.

In order to address the primary issue of concern--the stability of control strategy as affected by training duration and duration of breaks in training--pilot tests were designed to include breaks of varying duration within and between training sessions, and to give some subjects more extended experience in one training condition. The pilot tests thus allowed examination of changes in measured control strategy as a function of varying length breaks in training, as well as its development over longer periods of time. Also, the pilot tests permitted examination of the quality of HOPE model fits for various training-transfer combinations. At the end of the final session of pilot testing, each subject was interviewed concerning his strategy for performing the task. The interview was designed to help detect methodological problems in the design and to assess subjects' perceptions of control strategy.

The major results of the pilot tests can be summarized as follows. The results suggested that neither extended practice nor the duration of breaks in practice affect the stability of the measured control strategy. HOPE model fits to human behavior were acceptably good for the most part, and were also unaffected by these factors. The measured control strategy did appear to vary somewhat with changes in track frequency. On the whole, interview results were consistent with many of the theoretical constructs which form the basis for HOPE and for the measurement process used. Subjects reported variations in some aspects of control strategy not now permitted to vary in HOPE. Furthermore, they reported variation of control strategy in response to local conditions such as current error, current track demands, etc.

2. Validation Test Design

The pilot tests underlined the need to better understand the characteristics of control strategy. Given that need, a transfer task design was chosen for the validation tests. There were two groups of subjects. Each group of subjects was trained for five trials, each trial lasting three minutes with a one-minute break between trials. In their initial testing session, subjects completed one trial set (5 trials) in one training condition and then did a set of 4 trials in a second training condition. The next day, in their second testing session, subjects performed one trial in the second training condition and then returned to 5 trials in the first training condition. The subject groups are designated as Group ABA, or Group BAB, depending on whether they experienced the ¼ Hz track as their first and third

training condition (ABA) or the ½ Hz track as their first and third training condition (BAB). The experimental design is summarized in Table 3. It should be noted that subjects in the ABA group experienced mainly ¼ Hz track trials, so their data is the focus for examining CSP changes in this condition. Group BAB experienced mainly ½ Hz track trials, so their data is the focus for examining CSP changes in this condition.

TABLE 3

TRANSFER TASK PARADIGM FOR VALIDATION TESTING

	Trial Set (5 trials per set)		
Group	1[a]	2	3
ABA (12 subjects)	¼ Hz track, narrow guidelines	½ Hz track, narrow guidelines	¼ Hz track, narrow guidelines
BAB (12 subjects)	½ Hz track, narrow guidelines	¼ Hz track, narrow guidelines	½ Hz track, narrow guidelines

Note: On the first day of testing, each subject experienced 5 3-minute trials of Trial Set 1, then a 5-minute break followed by 4 3-minute trials of Trial Set 2. The second day, the final 3-minute trial of Trial Set 2 was presented, followed by 5 3-minute trials of Trial Set 3. Between-trial breaks were one minute in length.

[a]Tracks are random with cut-off frequency of ½ or ¼ Hz. Narrow guidelines are 1.6 cm on either side of the track.

The design allowed control strategy to be estimated for each training condition and at different points in learning. The obtained data could be used to examine the relationship between CSP estimates early and later in training, and between conditions. In addition, the relation between the control strategy estimates for the two different training conditions could be examined to determined whether the control strategy measured in the second condition could in any way be predicted from that measured for the first condition. The validation tests were supplemented by a post-experiment interview similar to that used in the pilot tests. Use of this interview was not based on an assumption that subjects have a completely accurate awareness of all aspects of their task learning and control strategy. However, pilot tests showed that subjects are able to speak in an apparently meaningful and informative fashion about

their experience. Their reported perceptions may contribute to the understanding of individual differences in control strategy.

Subjects were 24 paid volunteers from Air Force, Navy, and Army ROTC units on Campus. Since no sex differences in estimated control strategy had been observed in the preliminary tests and analyses, subjects of each sex were assigned at random to the two experimental conditions, with the restraint that neither group be exclusively male or female. One woman and eleven men were in Group ABA; one woman and eleven men were in Group BAB.

3. Apparatus

The apparatus for validation testing was that used in the preliminary testing and is displayed in Figure 6. The track and guidelines were presented on a Grinnell Systems GMR-27 digitally refreshed graphics display. The Conrac video monitor is 37.38 cm wide by 26.06 cm high. The track was generated by passing a pseudo-random signal through a low-pass filter. The track traveled downward from the top of the screen. About five seconds of track preview were available during tracking. The guidelines were centered around the track. Narrow guidelines appeared $\pm$ 1.6 cm directly horizontal of the track. Figures 7 through 10 show the tracks as they appeared in ¼ Hz and ½ Hz trial groups.

The subject controlled a cursor in the form of a small plus (+) visible on the screen. The cursor moves only in the horizontal dimension and is located halfway between the top and bottom of the video screen. Control was by means of a low friction isotonic stick with seven bits of position output, or 128 possible positions ($2^7 = 128$).

To minimize effects of past experience, the relationship between stick and cursor position, or control dynamics, was position type with non-linear first order lag. The nonlinearity was such that the cursor is more responsive to stick movement when the stick is moved in the middle of its range than at extremes. Learning these control dynamics was the fundamental learning task for the subjects and for the models.

A Perkin-Elmer Corporation mini-computer recorded control stick position every 40 msec. This same computer was used for data analvsis.

4. Procedure

Subjects were seated at a desk from which protruded the control stick and the key used to start each trial. About 1 m in front of them was the display screen. Instructions given to subjects on Day 1 and Day 2 are indicated below, and are fully consistent with the procedure used. Instructions given to subjects before Trial Set 1 on Day 1 were the following:

"This experiment involves using a control stick to move a small plus on the screen before you. When the experiment begins you will see a track moving down from the top of the screen. The track consists of a center line and two guidelines, one on either side of the center

Figure 6. Apparatus used in preliminary testing of HOPE.

Figure 7. Track in 1/4 Hz, narrow guideline condition.

Figure 8. Track in 1/4 Hz, wide guideline condition.

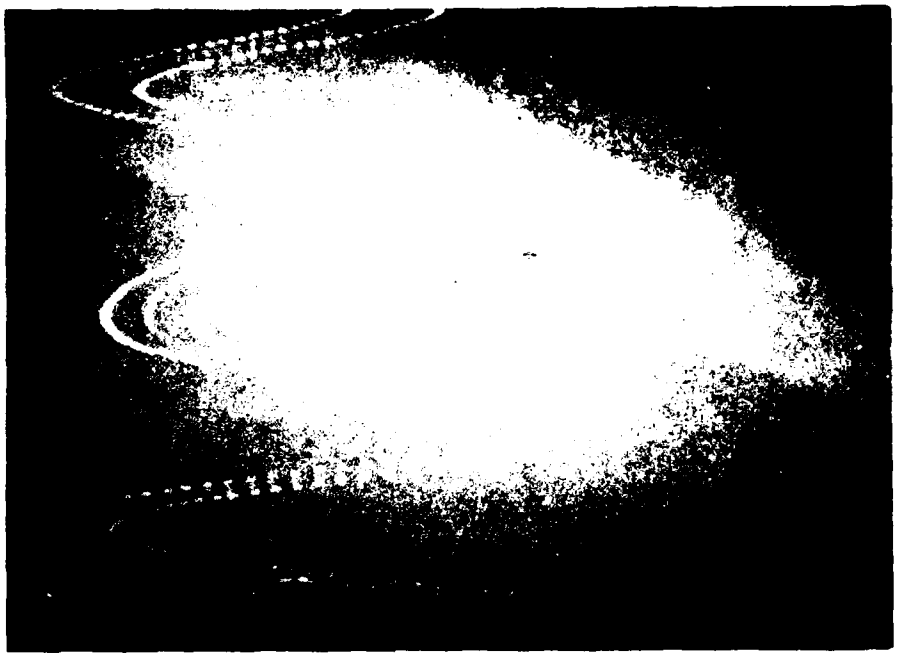


Figure 9. Track in 1/2 Hz, narrow guideline condition.

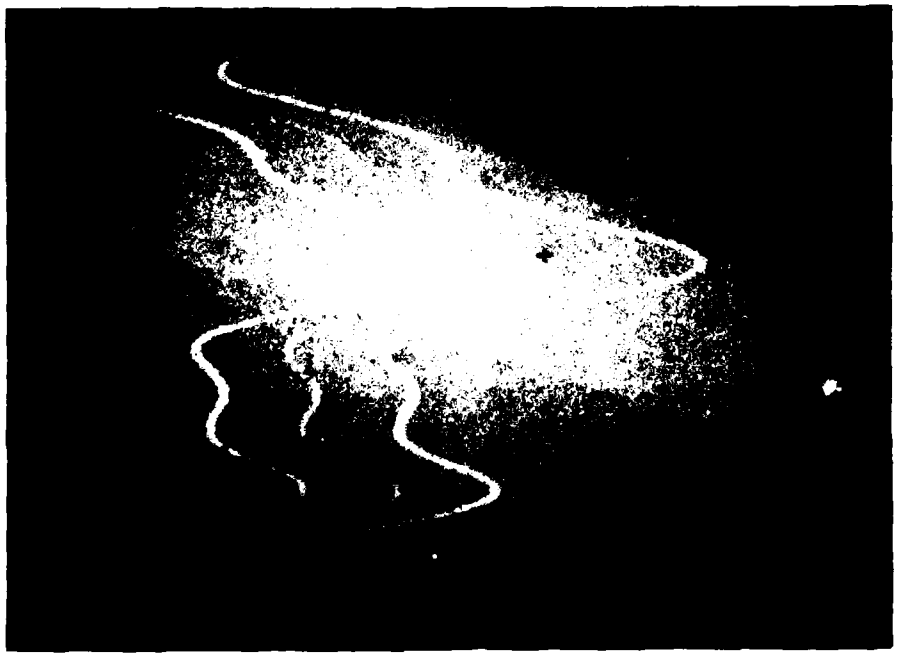


Figure 10. Track in 1/2 Hz, wide guideline condition.

line. Your task is to keep the plus on the center line and always within the guidelines. If you do cross the guidelines, please pay special attention to moving the plus back to the center line. Try to keep the plus on the center line and definitely within the guidelines.

"The experiment is divided into test trials, each lasting about three minutes. After each trial, your error will be displayed on the screen in inches. The displayed error is the total horizontal distance you were away from the center line for that trial. You will see the error on the trial just completed and the error on the previous trial so you can see if your performance is improving. The person who has the lowest overall error for the entire experiment will receive a ten dollar bonus.

"In the first half of today's session you will have five trials each lasting three minutes and each followed by a one-minute break. During the break, please look away from the screen to rest your eyes. Following the first four trials, you will have a five-minute rest period during which you may get up and walk around. Then you will have four more trials each followed by information about your error and by a one-minute break.

"Your first trial will begin soon. But first I'd like you to move the control stick back and forth to see how it feels. Please use your preferred hand to move the control stick. You will use your other hand to press the button which starts the trials.

"Further instructions will be presented on the screen. Please follow them very carefully and exactly. But, first, do you have any questions?

"Remember that your task is to keep the plus on the center line and definitely within the guidelines.

Now look at the screen and follow the instructions presented there."

Instructions given before Trial Set 2 on Day 1 were the following:

"You will now have four more trials, each followed by information about your error and by a one-minute break. Your task is to keep the plus on the center line and definitely within the guidelines. Remember that the person having the lowest total error will receive a $10 bonus.

"Further instructions will be presented on the screen. Please follow them very carefully and exactly.

"Any questions?

"Please begin."

On Day 2 the following instructions were given:

"Your task today is similar to that you performed yesterday. You will use the control stick to move the plus so as to keep it on the center of the track. The guidelines will again be present, and you should not allow the plus to go outside the guidelines. If you do cross the guidelines, please pay special attention to moving the plus back to the center of the track. Try to keep the plus on the center of the track, and definitely within the guidelines.

"Today you will have six trials, each lasting three minutes and each followed by information about your error. Remember that the person who has the lowest overall error will receive a $10 bonus. Please be sure to look away from the screen during the breaks between trials. At the end of today's session, I will explain the purpose of the experiment.

"Any questions?

"Now please follow very carefully the instructions on the screen."

As discussed in Section IV-B.2, an exit interview was conducted with each subject. The interview protocol is presented in Appendix A.

5. Choice of Best-fitting HOPE Models and Control Strategy Estimation Procedure

Each subject's control stick output was divided into 1350 time bins of 2 seconds each (15 trials x 180 sec)/2 sec = 1350 time bins, with 90 time bins per trial). Subject data was recorded every 40 msec (.04 sec) during testing, so there were 50 data points per time bin (2 sec/.04 sec = 50).

For each condition, HOPE was operated using all possible combinations of six values of COT (40, 80, 120, 160, 200, 240 msec), six values of ERRLIM (0, .6, 1.2, 1.8, 2.4, 3 cm), and six values of ADJUST (2, 4, 6, 8, 10, 12 control units). This resulted in 216 HOPE models (6 x 6 x 6 = 216) of human control stick position for each condition. For each time bin, a comparison was made between human control stick positions and the control stick positions for that time bin predicted by each of the 216 HOPE models. It is important to remember that human and model behaviors were compared only when each had equivalent amounts of experience in the task. For example, the behavior of a person in the fourth 2 sec bin of the first trial was not compared with models' behavior in the fourth 2 sec bin of the second trial, but only with model behaviors from the first trial. This restriction reflects the basic research assumption that when HOPE and humans have experienced the same proportion of the task the HOPE models' level of learning of external plant dynamics is the same as the human's.

A best-fit model for each time bin was selected, as described in the following paragraphs.

First, for each time bin the RMS difference between human output and the predictions of each of the 216 HOPE models was computed. The RMS difference calculation was performed in the following manner. Human control stick position was recorded every 40 msec throughout the experiment. HOPE generated 216 model predictions of control stick positions in each of the two training conditions. For each model in a given training condition, the position difference between model and human control stick position for each 40 msec measurement point was squared and summed within each 2 sec time bin. RMS difference is the square root of the sum.

Best-fit models for each time interval were selected by ranking models as to their goodness of fit with human behavior according to the RMS difference value. The model having the smallest RMS difference value was chosen as the best-fit model for that time bin. The final step in the procedure was to use best-fit models to infer human control strategy. Inferences were made in the following manner.

The CSPs of the best-fit model for each 2 sec time bin of human behavior were inferred to represent the control strategy used by the human in that 2 sec bin, if two conditions were satisfied.

- The best-fit model matched human behavior within a 10 percent criterion. To meet this requirement, RMS must be less than or equal to 12.8 screen units.

- The best-fit model was unique, as compared through the tenths places in the RMS difference score statistic. For example, if the first and second best-fit models had RMS differences with human output of 8.9, no estimate of control strategy was made for that time interval.

Thus, only models which matched human behavior very closely and which were also uniquely best, at an appropriate accuracy level, were used to infer human control strategy.

C. Results and Discussion

The results are presented for each research question, as compared with predictions made in Section IV-B. Results from the post-experiment interview follow these. A summary discussion completes the Section. Certain issues raised by the results will be discussed more fully in the assessment of the total research approach, which is provided in Section V.

1. Do HOPE Models Match Human Behavior Well Enough to Permit Estimation of Human Control Strategy?

In previous discussion, the standards by which the quality of matching should be judged were set forth as follows: First, HOPE models should match within 10 percent of the control stick range of motion (within 12.8 screen units) at least 80 percent of the time (for 1080 of 1350 2 sec time intervals) for 90 percent of the subjects (22 subjects). Second, inferences would not be made even from physically

close matches, if there existed no uniquely best match, as measured through the tenths place in the RMS difference score. Third, HOPE models should match equally well human behavior in varying conditions of learning and training.

Data bearing on these three issues is presented in Tables 4 and 5. It is clear from these tables that HOPE models meet the quality and uniqueness criteria for inference that were set prior to the analysis. Note that use of the uniqueness criterion reduced the percentage of time bins matched more than did the quality of matching criterion, a result which will be discussed in Section V-C.1.a of this report. Further, there seems little difference in the percent of behavior matched within the criterion whether ½ Hz or ¼ Hz cutoff frequency tracks were followed. About 85 percent of the time bins of human behavior were acceptably and uniquely matched in both conditions.

As is evident from the RMS difference averages in both Table 4 and 5, however, matching of ¼ Hz tracking behavior was, on the average, about twice as good as matching of ½ Hz tracking behavior. Furthermore, the average quality of matching ¼ Hz behavior in Group ABA appeared to improve over the course of training. These differences will be discussed in the model assessment section (Section V-B.1.a).

For purposes of these analyses, it can be said that the preset criterion for useful inferences of control strategy was met about the same percent of the time for both conditions and track frequencies. This fact means that comparisons between the control strategy parameters estimated in both track frequencies can be made.

2. Does Estimated Control Strategy Vary over the Course of Task Learning?

It was expected, prior to testing, that control strategy would become task-specific strategy as skill in the task developed. Indeed, an important contributor to skill development may very well be change in control strategy. If this is so, then control strategy should change when performance improves. Predictions were made of a downward trend in COT, especially in ½ Hz trials experienced by Group BAB; and of change (direction unspecified) in both ADJUST and ERRLIM. The changes for ADJUST and ERRLIM were expected in either BAB or ABA, but not necessarily both. See Part A.1 of this Section for justifications of these predictions. Testing of these predictions was conducted separately for each of the groups, and will be discussed in that fashion.

a. Group ABA results--These subjects tracked for 15 trials in the ABA design. Their A trials were of a ¼ Hz cutoff frequency random track; their B, a ½ Hz track. Figure 11, 12, 13 and 14 show the mean RMS error, COT, ERRLIM, and ADJUST over time for this group. Table 6 shows the average change in each variable between the first and last trial. Considering just the RMS error for the ABA group in their ¼ Hz trials, it is apparent that, on the average, subjects improved between Trial 1 and Trial 15. Indeed, the average change in RMS error between Trial 1 and Trial 15 was significantly different from zero, as is indicated in Table 6 ($t(11) = 12.2$, $p < .001$, one tailed). According to a conventional measure of skill, skill development occurred.

TABLE 4

HOPE MATCHING QUALITY FOR SUBJECT GROUP ABA (N = 12)

Training Condition	Average Percent of Time Bins Matched		No. of Subjects Matching within Criteria	RMS Difference Averages Trial 1	2	3	4	5
	within 10%	99.9	12					
¼ Hz	also uniquely	82	11	3.3	3.1	2.9	2.8	2.7
	within 10%	95	12					
½ Hz	also uniquely	85	11	6.1	6.4	6.3	6.3	6.4
	within 10%	99.9	12					
¼ Hz	also uniquely	83	12	3.3	2.9	2.8	2.7	2.6
	within 10%	98	12					
TOTAL	also uniquely	83	11					

TABLE 5

HOPE MATCHING QUALITY FOR SUBJECT GROUP BAB (N = 12)

Training Condition	Average Percent of Time Bins Matched		No. of Subjects Matching within Criteria	RMS Difference Averages Trial 1	2	3	4	5
	within 10%	94	12					
½ Hz	also uniquely	86	12	6.6	6.5	6.6	6.4	6.4
	within 10%	100	12					
¼ Hz	also uniquely	84	12	3.5	3.2	2.9	3.0	3.1
	within 10%	96	12					
½ Hz	also uniquely	86	11	6.1	6.2	6.2	6.2	6.2
	within 10%	97	12					
TOTAL	also uniquely	85	11					

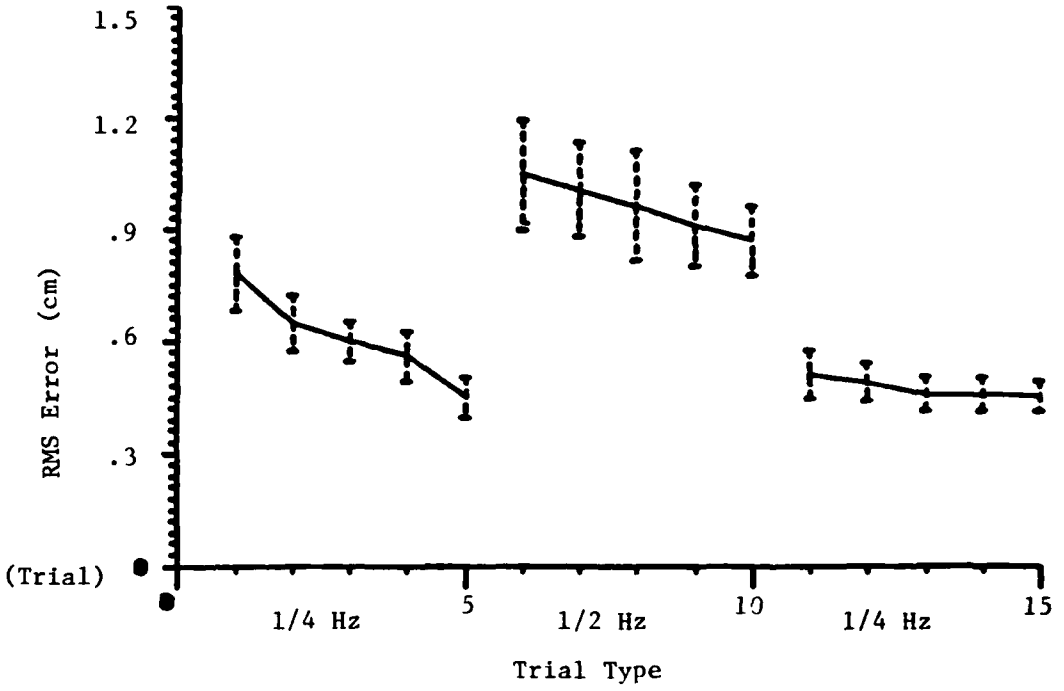


Figure 11. Mean RMS Error for Group ABA over Trials

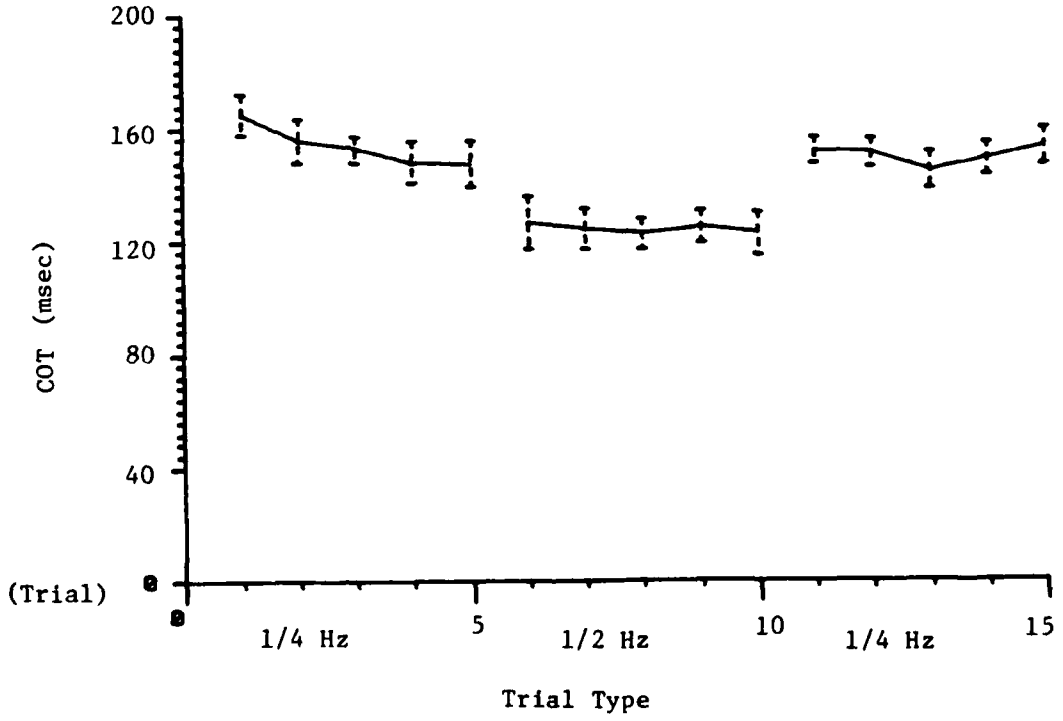


Figure 12. Mean COT Estimates for Group ABA over Trials

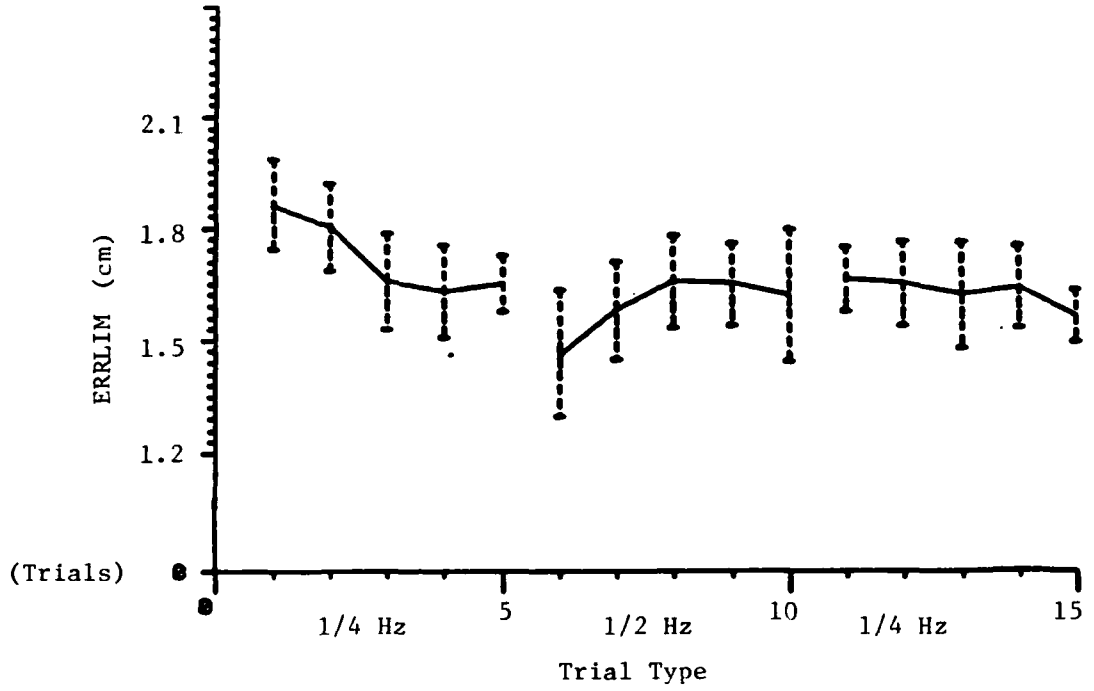


Figure 13. Mean ERRLIM Estimates for Group ABA over Trials

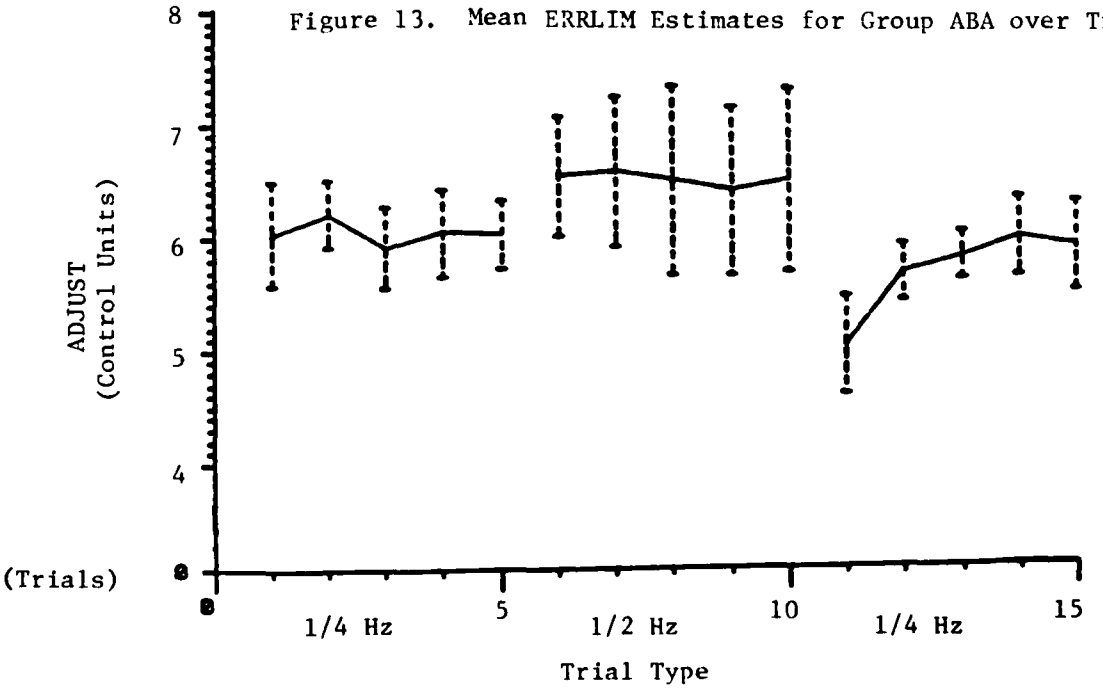


Figure 14. Mean ADJUST Estimates for Group ABA Over Trials

TABLE 6

MEAN CHANGE IN CONTROL STRATEGY AND ERROR BETWEEN FIRST AND LAST TRIAL

ERROR		COT[a]	ERRLIM[b]	ADJUST[c]	RMS[d]
Group ABA (N = 12)	d	=.31*	.95**	.10	1.10***
	s_m	= .08	.16	.13	.09
Group BAB (N = 12)	d	= 1.06**	-.32	-.59*	1.75***
	s_m	= .09	.18	.18	.31

Note: d is mean change
s_m is the standard error of the mean

[a]One COT unit = 40 msec
[b]One ERRLIM unit = .294 cm
[c]One ADJUST unit = .8% of total control range
[d]One RMS error unit = .294 cm

*p < .005, one-tailed
**p < .002, two-tailed
***p < .001, one-tailed

Was there a corresponding change in control strategy used by subjects in this condition? Figures 12 and 13 show the time course of the group averages of the control strategy parameters for which individuals in Group ABA changed significantly over the course of learning between Trial 1 and Trial 15--COT and ERRLIM. Considering, again, just Group ABA, in Table 6, individual measures of ERRLIM decreased, on the average, between Trial 1 and 15, and the average difference was significantly different from zero ($t(11) = 5.94$, $p < .002$, two-tailed). COT, as is indicated in Figure 12, also showed a tendency to decrease between the first and last trial. Table 6 shows the size of that change was small, but significant ($t(11) = 3.88$; $p < .005$; one-tailed). Thus, experience in the relatively easy ½ Hz trial groups was associated with significant performance improvement as well as significant decrease in the estimate for the internal standard for performance, ERRLIM. COT also reduced somewhat. ADJUST showed no developmental pattern.

b. Group BAB results--These subjects tracked ½ Hz cutoff frequency tracks as their two B groups, ¼ Hz tracks as their A group. Figure 15 shows the average RMS error for the 12 subjects in Group BAB. Considering just the ½ Hz trials, this conventional measure of skill shows an apparent increase. Table 6 confirms this. The average decrease in RMS error, between Trial 1 and 15, was significantly greater than zero ($t(11) = 5.65$, $p < .001$, one-tailed). Were there comparable changes in control strategy parameters measured? Figure 16 shows estimates of COT over trials. As the ½ Hz group mean trend in Figure 16 suggests, the average individual change from Trial 1 to 15 was significantly different from zero ($t(11) = 11.77$, $p < .001$, two-tailed). Figure 18 shows the group average ADJUST measures for the Group BAB subjects. Eleven of 12 subjects increased between Trial 1 and 15, and Table 6 reveals that the average increase was significantly larger than zero ($t(11) = 3.28$, $p < .005$, one-tailed). No consistent individual changes in ERRLIM occurred, as is suggested by the lack of trend in the group mean shown in Figure 17.

c. Consistency of results--These results from the analyses of measures taken in Group ABA and Group BAB are consistent with the predictions made previously, in that COT reduced; ADJUST changed so that group BAB measures were larger than ABA measures; and ERRLIM showed a reduction in Group ABA, so that ABA ERRLIM is the smaller of the two groups. The results are also consistent with other results obtained in these tests, which will be discussed later in Section IV-C.6.

3. Does Control Strategy, as Measured by HOPE, Reflect Differences in Training Conditions?

It was expected prior to testing, that the frequency manipulation used in these validation tests would be effective in producing different control strategies in the two groups. Group ABA experienced their first and last trial sets in ¼ Hz conditions, and Group BAB their first and last trial sets in ½ Hz conditions. Though each experienced the alternate frequency in the middle trial set, the different lengths of experience in each suggested that by the last trial (15), there should be control strategy differences between the two groups specific to the differing frequencies of the first and last trial sets. In particular, it was expected that COT would be lower in ½ Hz than in ¼ Hz conditions; ERRLIM would be larger in ½ Hz than in ¼ Hz conditions; and ADJUST would be larger in ½ Hz than in ¼ Hz conditions. The differences were expected to be significant by the last trial, but not on the first trial. The reversal of tracking conditions during the middle trial set also permitted comparisons between control strategies for ¼ and ½ Hz tracks, developed after prior experience in the alternate condition.

These hypotheses were largely supported in the following ways. As can be seen in Table 7, even during the first trial, COT was significantly shorter ($t(22) = 3.38$; $p < .005$, one-tailed) for the BAB group than the ABA group. Although first trial differences had not been predicted, in retrospect the result seems reasonable. COT is not related to the variable lag--the distinctive aspect of these control dynamics--but is rather obviously related to frequency. Subjects do have tracking

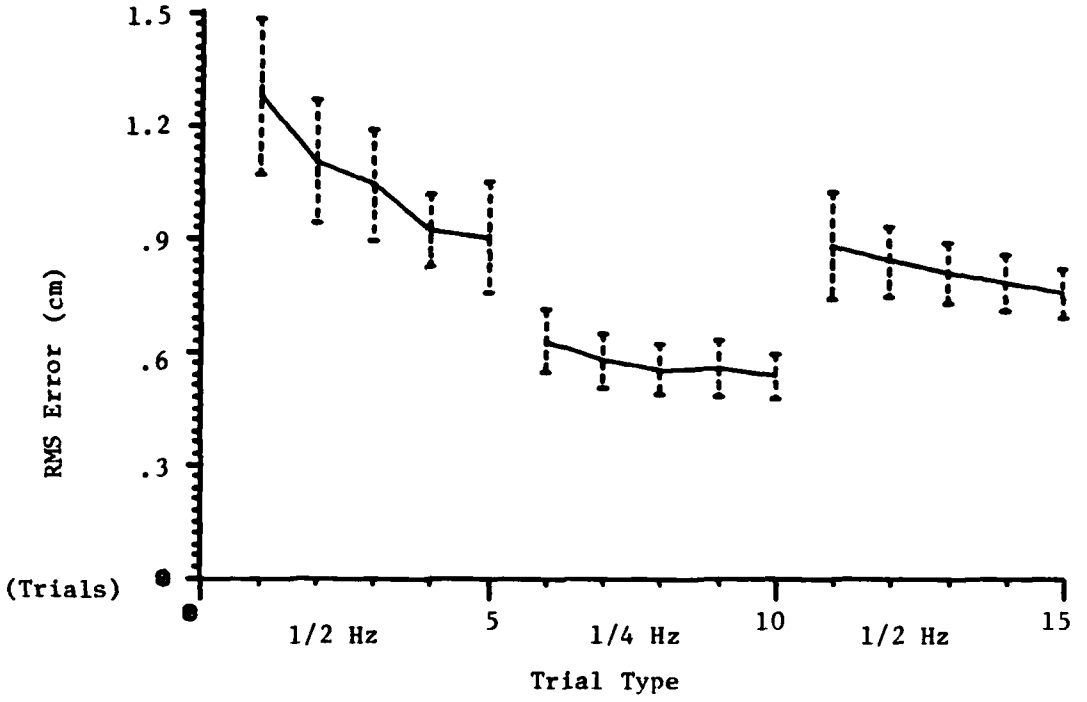


Figure 15. Mean RMS Error for Group BAB over Trials

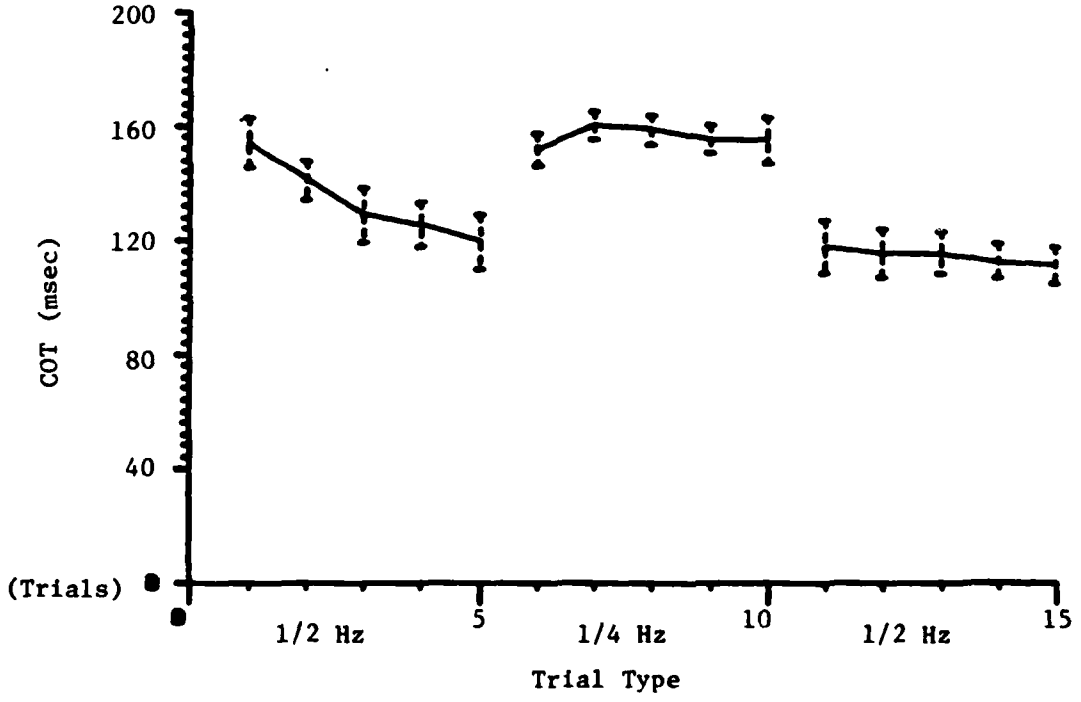


Figure 16. Mean COT Estimates for Group BAB over Trials

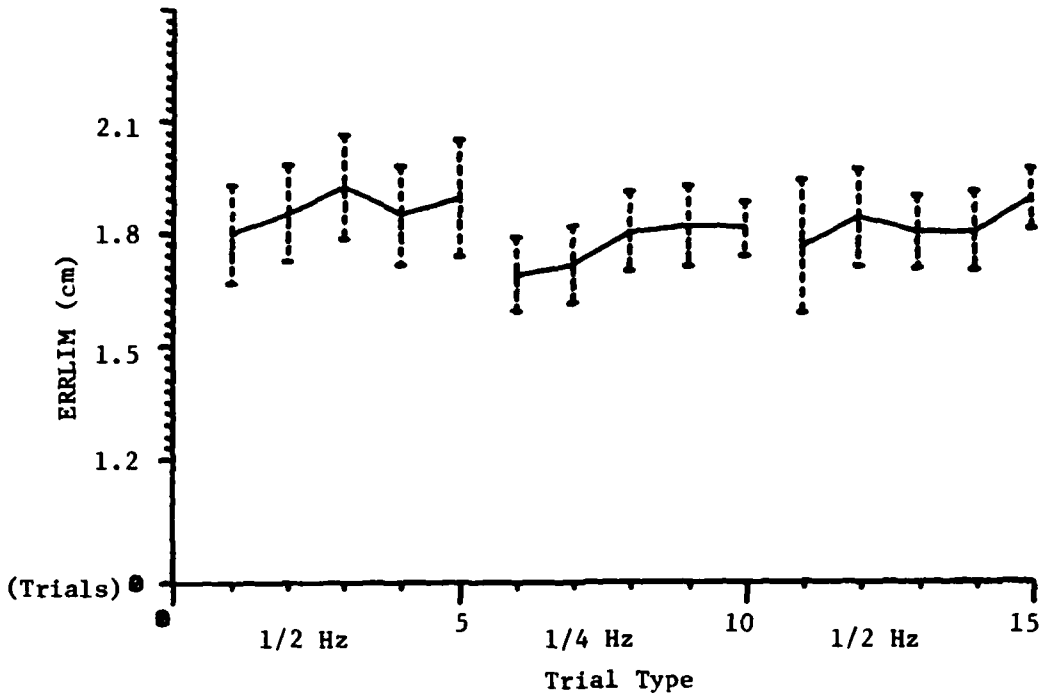


Figure 17. Mean ERRLIM Estimates for Group BAB over Trials

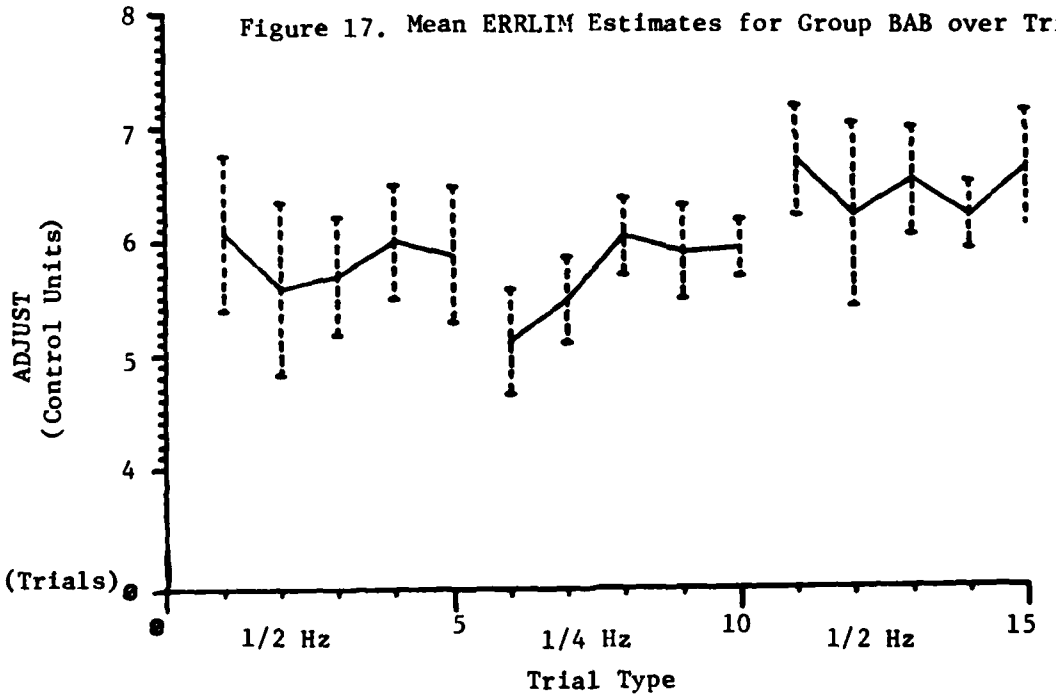


Figure 18. Mean ADJUST Estimates for Group BAB over Trials

TABLE 7

CSP MEANS FOR EACH GROUP, BETWEEN GROUP DIFFERENCES FOR TRIALS 1, 15, AND 10

Trial	Subject Group	Track Frequency	CSP		
			COT[a]	ERRLIM[b]	ADJUST[c]
1	ABA	¼ Hz	4.12	6.20	6.04
	BAB	½ Hz	3.85	5.99	6.07
	Difference		.27*	.21	- .03
15	ABA	¼ Hz	3.82	5.21	5.96
	BAB	½ Hz	2.78	6.30	6.62
	Difference		1.04**	-1.09**	.66**
10	BAB	¼ Hz	3.88	6.03[d]	5.93[d]
	ABA	½ Hz	3.06	5.39	6.53
	Difference		.82**	.64	- .60

[a]One COT unit = 40 msec
[b]One ERRLIM unit = .294 cm
[c]One ADJUST unit = .8% of total control range
[d]Variance of these subject group means were too different for comparisons using t distribution.

*p < .005; one-tailed
**p < .001; one-tailed

experience and might be expected to discover this task-related aspect of control strategy rather quickly. COT was also significantly shorter, as expected, for the 15th trial of the BAB Group (a ½ Hz track trial) than for the 15th trial of the ABA Group (a ¼ Hz track trial) ($t(22) = 15.07$; $p < .001$, one tailed). The differences in COT apparently became more exaggerated with experience in the task.

Another difference in control strategy that was predicted was for ERRLIM, the parameter thought to represent an internal performance standard. On the last training trial, but not the first, ERRLIM in the ¼ Hz A group was significantly smaller than in the ½ Hz B group ($t(22) = 10.38$, $p < .001$, two-tailed). The result is by no means an unreasonable one, since the ¼ Hz condition is so much easier, as measured by RMS error, than the ½ Hz condition. An easier task, at least prior to the point where boredom sets in, should permit stricter internal standards than a more difficult one. In this case, subjects had been told a $10 bonus prize would be awarded the individual (in each group) with the lowest overall error. There were no indications in performance or in the post test interview that subjects had become bored. Therefore, the internal performance standard should have been different in the two groups.

The other aspect of control strategy which was expected to vary with conditions, and was also expected to require some experience for its development was ADJUST. Here the a priori expectations were also borne out. There were no significant differences in ADJUST on Trial 1. By the last trial, however, subjects in the BAB group had significantly larger estimated ADJUST values ($t(22) = 3.53$; $p < .001$, one-tailed).

The same comparisons were made between group means in the last trial of the middle set of trials, trial 10. Thus, for example, the ABA group of subjects which had a first set of five trials of ¼ Hz track switched to five of the ½ Hz, for their second trial set. In the absence of carryover effects, second set CSP estimates might be expected to follow the same condition-related patterns observed in the other sets. Several factors make this a weak assumption, however. First, there is no clear reason to assume the absence of carryover effects. Second, five trials might not be sufficient for a new control strategy to develop. Third, the variances of CSP estimates for two of the parameters in the second trial set were significantly different between groups. ABA variances for ERRLIM and ADJUST were much larger than is consistent with an equal variance assumption ($F(11,11) = 5.8$; $p < .01$). For COT, however, variance of estimates were apparently similar ($F(11,11) = 1$; $p < .05$). The ABA group, after experiencing 5 trials of the ½ Hz track, had significantly smaller COT than the BAB group after their 5 trials in the ¼ Hz track ($t(22) = -10.25$; $p < .001$, one-tailed).

4. Do Measures of Control Strategy Show Predictive Validity?

It was hypothesized, prior to testing, that several types of predictive relationship might be demonstrated for measures of control strategy made using HOPE. First, control strategy measured early in learning

might be expected to be predictive of control strategy measured late in learning. Second, control strategy measured in one training condition might be predictive of control strategy measured in another setting. Third, control strategy measures might show relationship to errors made at the same or later periods in the same task setting. Fourth, control strategy measures from one task setting might predict errors made in a second task setting. Finally, selection of a representative control strategy for an operator might permit predictions of later behavior and performance in the same or a different task.

Each of these hypotheses was tested in a preliminary manner. The term preliminary is appropriate because only the simplest level of predictive relationship was tested. The linear correlation was used to examine the first four of these predictive relationships, because the range of values used was too narrow to permit investigation of curvilinear trends. Relationships among the parameters (e.g., COT and ERRLIM) were not explored in these analyses either, due to the limited scope of these tests. Results of these investigations are discussed below.

Table 8 illustrates the extent of the linear relationship found between control strategy measured in trial one or five and control strategy measured in the last trial of tracking in the same condition. (Trial averages for individual subjects were used in these calculations). For example, for Group ABA, Table 8 shows the relationship between COT on Trial 1 (a ¼ Hz track trial) and COT on Trial 15 (also a ¼ Hz track trial). In this case, the correlation was .222.

As is shown in Table 8, only COT and ERRLIM in the fifth trial showed a statistically significant relationship to the average COT and ERRLIM measured in Trial 15 (for COT, $t(22) = 3.04$; $p < .005$, one-tailed: for ERRLIM, $t(22) = 2.32$, $p < .025$, one-tailed). These are both positive linear relationships, as predicted. Both COT and ERRLIM tend to be larger after 45 minutes of tracking if they were larger after 15 minutes of tracking. Trial 1 measures are not at all predictive of Trial 15 measures; and ADJUST on Trial 15 shows no relationship to its Trial 5 measure, either. The absence of ADJUST relationships was not expected.

Table 9 portrays the extent of the linear relationship between like control strategy measures taken in two different frequency tracking tasks. These data were calculated by pairing a measure from the last trial in a group of trials in one training condition with a measure of a like CSP from either the first or the fifth trial in a different training condition. Table 9 is, then, one way to consider the extent of carryover effects resulting from changing track frequencies. As the table shows, COT measured in the last ½ Hz track trial in a trial set is significantly related to the COT used in the first ¼ Hz trial of the next trial set, but not to the fifth ¼ Hz trial of the next trial set ($t(22) = 3.5$, $p < .005$, one-tailed). ERRLIM on the last trial of a ½ Hz trial set is related positively to the ERRLIM in the fifth trial of a following ¼ Hz trial set ($t(22) = 5.14$; $p < .001$, one-tailed), but not to that used in the first trial of the following ¼ Hz trial set. Neither ADJUST in ½ to ¼ Hz transfer nor any of the

TABLE 8

PREDICTIVE VALIDITY: LINEAR RELATIONSHIP BETWEEN LIKE CSP's IN EARLY(X) AND LAST(Y) TRIALS IN SAME FREQUENCY TRACKING TASK

	CSP(Y)		
Trial of X Measure[a]	COT	ERRLIM	ADJUST
One			
r_{XY}	.222	-.286	.399
$b_{Y \cdot X}$	.451	-.541	.273
Five			
r_{XY}	.545**	.444*	-.024
$b_{Y \cdot X}$	.608	.444	-.013

Note: All Y measures were taken in 15th trial of tracking, after approximately 42 mins of experience.

[a] r_{XY} is the linear correlation between the predictor X and the criterion Y.
$b_{Y \cdot X}$ is the regression coefficient for the best-fitting straight line predicting Y from X.

*p < .05, two-tailed
**p < .01, two-tailed

TABLE 9

PREDICTIVE VALIDITY: LINEAR RELATIONSHIP BETWEEN LIKE CSPs
IN TWO DIFFERENT FREQUENCY TRACKING TASKS

Trial Type for X Measure[a]		CSP					
		COT		ERRLIM		ADJUST	
		¼ Hz Next Trial(Y)	¼ Hz 5th Trial(Y)	¼ Hz Next Trial(Y)	¼ Hz 5th Trial(Y)	¼ Hz Next Trial(Y)	¼ Hz 5th Trial(Y)
½ Hz	r_{XY}	.598*	.159	.287	.739**	.241	.323
	$b_{Y \cdot X}$	.466	.108	.174	.642	.134	.166
		½ Hz Next Trial(Y)	½ Hz 5th Trial(Y)	½ Hz Next Trial(Y)	½ Hz 5th Trial(Y)	½ Hz Next Trial(Y)	½ Hz 5th Trial(Y)
¼ Hz	r_{XY}	.090	.018	-.007	.118	.141	-.002
	$b_{Y \cdot X}$	.120	.017	-.006	.103	.173	-.003

r_{XY} is the linear correlation between the predictor X and the criterion Y.
$b_{Y.X}$ is the regression coefficient for the best-fitting straight line predicting Y from X.

*p < .01, two-tailed
**p < .002, two-tailed

parameters in the ¼ to ½ Hz transfer showed significant linear relationship. The directions of the two relationships observed were as predicted. Again, the absence of ADJUST relations was not predicted, nor was the absence of ¼ Hz to ½ Hz transfer effects.

Table 10 displays the extent of linear relationships between control strategy and either concurrent or later RMS position error. Since a control strategy is assumed to be task-specific, in fact, tailored for good performance, correlations were calculated separately for data taken in each track frequency. For example, ½ Hz training condition relationships between CSPs and RMS error were evaluated as follows. CSPs measured in the fourth trial of ½ Hz condition trial sets were correlated with RMS error measured for that trial and with RMS error measured for the following trial.

As Table 10 shows, consistent and significant relationships exist in ½ Hz tracking between the control strategies estimated for human subjects and their concurrent and future performance. COT is positively related to error, both in same and in the next trial after measurement (same trial $t(34) = 4.69$, $p < .001$, one-tailed; next trial $t(34) = 3.96$, $p < .001$, one-tailed). Thus, in ½ Hz tracking, use of a smaller COT is associated with better performance, as predicted.

ERRLIM, in contrast, is related negatively to performance, both in same and in next trial after measurement (same trial $t(34) = -2.93$, $p < .005$, one-tailed); next trial $t(34) = -2.76$, $p < .005$, one-tailed). In ½ Hz tracking, lower error is associated with a larger ERRLIM. This relationship is opposite to that expected. ADJUST, as well, in ½ Hz conditions, is related negatively to performance (same trial $t(34) = -2.56$, $p < .01$, one-tailed; next trial $(t)34 = -2.414$, $p < .025$, one-tailed). Subjects estimated to be using higher values of ADJUST had lower tracking error, as had been predicted.

In ¼ Hz tracking, only COT showed any relationship to error (same trial $t(34) = 2.247$, $p < .025$, one-tailed). No other relationships with error were significant, in ¼ Hz tracking. The absence of ¼ Hz predictions using ERRLIM and ADJUST was not predicted, nor was the presence of the COT relationship.

Thus, in ½ Hz tracking, good performance is associated with high frequency, perhaps slightly jerky, movements, and with a relatively lenient internal performance standard. Performance in ¼ Hz tracking is associated only with relatively high frequency movements.

A fourth type of linear relationship tested involved the ability of CSP measures taken prior to transfer to predict error measures taken soon after transfer. Table 11 shows the modest relationship revealed when CSP measures from the last trial prior to transfer from one frequency to another are used to predict RMS error measured in the first trial of the new condition. Only ½ Hz measures were predictive, and the

TABLE 10

CONCURRENT AND PREDICTIVE VALIDITY: LINEAR RELATIONSHIP BETWEEN CONTROL STRATEGY AND RMS ERROR

a) ½ Hz Track: Relationships Between CSPs and RMS Error

Trial of Predicted RMS (Y)		CSP (Predictors) COT(X)	ERRLIM(X)	ADJUST(X)
Same	r_{XY}	.626***	-.450**	-.402**
	$b_{Y \cdot X}$	1.1	-.339	-.258
Next	r_{XY}	.563***	-.429***	-.382*
	$b_{Y \cdot X}$	1.059	-.347	-.263

b) ¼ Hz Track: Relationships Between CSPs and RMS Error

Trial of Predicted RMS (Y)		CSP (Predictors) COT(X)	ERRLIM(X)	ADJUST(X)
Same	r_{XY}	.360*	.105	.107
	$b_{Y \cdot X}$	.534	.062	.072
Next	r_{XY}	.151	.199	.024
	$b_{Y \cdot X}$	.188	.099	.014

[a] r_{XY} is the linear correlation between the predictor X and the criterion Y.
$b_{Y \cdot X}$ is the regression coefficient for the best-fitting straight line predicting Y from X.

*p < .025
**p < .01
***p < .002

TABLE 11

CORRELATIONS OF CSPs WITH FUTURE ERROR AFTER TRANSFER

Transfer Type		CSPs (Predictors) COT(X)	ERRLIM(X)	ADJUST(X)
½ Hz to ¼ Hz				
	r_{XY}	.484*	.097	-.389**
	$b_{Y \cdot X}$	.567	.036	-.123
¼ Hz to ½ Hz				
	r_{XY}	.167	.098	.273
	$b_{Y \cdot X}$	.304	.101	.272

Note: RMS Error (Y) was measured in the first trial after transfer, for both transfer types.

*p < .05, one-tailed
**p < .01, one-tailed

evident associations are quite modest. COT used in ½ Hz is positively associated with error after transfer to ¼ Hz tracking ($t(22) = 2.59$, $p < .01$, one-tailed). That is, persons using higher COTs in ½ Hz tracking were likely to have higher error after transfer than those using lower values. Subjects using higher ADJUST values in ½ Hz may show lower error after transfer to the ¼ Hz conditions than those using lower values ($t(22) = -1.97$, $p < .05$, one-tailed). The direction of these relationships was as predicted. The absence of ½ Hz ERRLIM as a predictor, as well as the absence of predictive power from measures taken in ¼ Hz tracking were not predicted.

Quite a different approach was taken in the final phase of the examination of criterion-related validity of HOPE measures. This approach involved a representative model approach discussed earlier. The procedure followed was this. In the fifth trial of training, the tabulation of frequencies of selection for each CSP was used to select a modal value for each CSP, as a representative control strategy, one from each subject group. In addition, mean values for CSPs were used to select a second candidate representative control strategy for each group. The fifth trial was selected for the choice of CSP values,

because it was believed that control strategy would be more likely to be task-related after some experience than initially. These four representative models--two from each subject group--were then run through the 15 trials of the experiment, in the design used by the subject group of which they were representative. The acceptability of these representative models was judged by one criterion; the number of 2 sec bins for trials after that in which the selection was made in which they matched each subject's behavior within 12.8 (the quality of matching criterion used earlier).

It was predicted that the models selected by this means would match half or more of the subjects in the group to within the quality criterion of 12.8, at least 80% of the time. The similarity of subjects' last trial control strategies suggested that this should be so. Table 12 displays the results from this ad hoc assessment of the power of a single HOPE model to predict behavior in the remaining ten trials of testing. The model chosen on the basis of modal values in Group BAB is clearly superior to the others. Not only did that model produce behaviors acceptably close to all subjects in Group BAB, in their ¼ Hz trials, but also produced behaviors acceptably close to 67% (8) of subjects in Group BAB in the more difficult-to-match ½ Hz trials.

No other model came so close to ½ Hz behavior, though all four did well matching ¼ Hz behavior.

A puzzling aspect of these results is the fact that four distinct models, with distinct CSPs--i.e., COTs of 2, 3, 4, 5; ERRLIMs of 6, 8, 10; and ADJUSTs of 2, 6--should match ¼ Hz behavior of all subjects so well. The matching quality is, however, not nearly so good as that of best-fit models. As an illustration of this, see Table 13, in which the trial by trial mean RMS differences between one subject and that subject's best-fit model are compared to the RMS differences between that subject and representative models. These data are typical for all subjects, and demonstrate that although representative models predict well, in ¼ Hz tracking, they do not predict nearly so well as do models representing the time-varying control strategies of the subject.

The matching produced for ½ Hz trials was much poorer, with only one of the four models acceptably matching the behavior of half or more of the subject group it represented. The model which accomplished this was selected from modal values in the fifth trial of ½ Hz tracking for the BAB group.

These modest results are an indication, at least, that the psychologically-based simulation HOPE may, in the future, produce accurate predictions of the details of behavior in continuous control. Such ability to predict the details of behavior in response to a changing environment is the unique property of a simulation approach, compared to other measures of control strategy. The utility of such predictions can be realized only after the simulation HOPE has included in it a representation of the control strategy development process, so that its behavior reflects not only increasing knowledge of the control dynamics of the external plant, but also an increasingly appropriate control strategy.

TABLE 12

MATCHING QUALITY FOR REPRESENTATIVE MODELS

Trial Types	Selection Basis for CSPs: Mean	Selection Basis for CSPs: Mode
	CSP Values[a, b]	
¼ Hz		
ABA Group	4, 6, 6	5, 8, 2
Mean % Bins Matched Within 12.8	95	97
% Subjects < 12.8 for 80% or more Bins	100	100
BAB Group	3, 6, 6	2, 10, 2
Mean % Bins Matched Within 12.8	97	99
% Subjects < 12.8 for 80% or more Bins	100	100
½ Hz	CSP Values	
ABA Group	4, 6, 6	5, 8, 2
Mean % Bins Matched Within 12.8	67	60
% Subjects < 12.8 for 80% or more Bins	8	0
BAB Group	3, 6, 6	2, 10, 2
Mean % Bins Matched Within 12.8	73	81
% Subjects < 12.8 for 80% or more Bins	17	67

[a]The order of the parameters is COT, ERRLIM, ADJUST.

[b]CSP values are expressed in units representing the following quantitites:
1 COT unit = 40 msec
1 ERRLIM unit = .294 cm
1 ADJUST unit = .8% of control stick range

TABLE 13

COMPARISON OF MEAN RMS DIFFERENCES FOR BEST FITTING AND REPRESENTATIVE MODELS FOR ONE SUBJECT

RMS Differences Between Human and	Trial and Track Type									
	½ Hz					¼ Hz				
	6	7	8	9	10	11	12	13	14	15
Best-Match Model	5.3	5.8	5.3	5.8	5.7	3.1	2.9	2.8	2.6	2.7
Mean Representative	8.3	8.8	8.6	9.0	8.3	6.6	5.9	6.1	6.0	5.6
Mode Representative	8.1	8.2	8.6	8.5	8.8	5.7	5.4	5.3	5.1	4.9

Note: RMS values in this table were calculated in terms of units of control, where 1 unit of control represents .8% of total range.

5. Results from the Post-Experiment Interview

As an additional check on the validity of the construct of control strategy, all subjects who participated in the validation tests were interviewed about aspects of their test experience which related to control strategy. In the discussion below, results from this interview are described in a general fashion. Details of responding are presented in Appendix B.

First, subjects were asked what they had done to try to improve performance. Their responses fell into the following categories:

a) changed style of control stick manipulation (e.g., tried not to over-correct, changed hand positions, tried to move smoothly),

b) learned plant dynamics (e.g., learned sensitivity between stick and plus sign),

c) changed level of concentration (e.g., concentrate more),

d) used memory (e.g., remember road patterns), and

e) changed what was attended to (e.g., would preview for a moment then concentrate on a few dots in front of plus, then repeat, learned to focus on center line).

Notice that in these responses, aspects of both control strategy (Items a, d, e) and of learning of the internal plant model (Item b) were mentioned by some subjects.

A second question asked focused more specifically on what was learned by subjects. They were asked what they thought they learned that helped them improve their performance. This question drew responses in categories much like the previous categories:

a) learned control dynamics (e.g., learned how far to push stick for certain types of 'S' curves, learned lag in stick response),

b) learned style of control stick manipulation (e.g., hold hand steady, find best grip),

c) learned tracks (e.g., learned road patterns, learned where roads began, learned there were two roads),

d) learned effective level of attention/arousal (e.g., concentrate, be relaxed), and

e) learned good attention strategy (e.g., learned to ignore some preview, used dot spacing to cue speed of bat handle required).

This reported learning includes internal model development (Item a), input predictor development (Item c) and development of input (Item e) and output (Item b) aspects of control strategy. Notice that input predictor development was not thought to be necessary for preview tracking, yet subjects reported it as a learning which related to their performance improvement.

Because the guidelines were defined to the subjects in much the way ERRLIM is defined in the model, it seemed appropriate to try to understand how subjects perceived them. They were asked if they paid attention to the guidelines. A slight majority of responses indicated that the guidelines were attended-to. They were frequently used as a measure of error, and as an occasional definition of acceptable performance by subjects who reported they tried to stay within guidelines on sharp turns, rather than just staying on the center line. They were also reportedly used as a visual cue for the shape of the upcoming track. A substantial minority of subjects reported that they ignored them.

Three questions in the interview focused on the subject-reported variation of aspects of control strategy represented in HOPE. For instance, they were asked if they varied how often they picked new positions for the control stick. Many subjects reported variation

in this aspect. Some reported increasing frequency with time, some reported decreasing with time. Many reported increasing frequency of moves when the track curved sharply; some reported decreasing frequency in such curves. A few reported no change. This question was meant, of course, to correspond to COT. The answers appear to correspond, in many cases, to a variable COT, which is responsive to track frequency and to experience.

Subjects were also asked if they varied the amount of deviation from the center that was judged as acceptable. Many reported that as the testing progressed, a decreasing amount of distance from the center was acceptable. Also, some subjects said the "slow road" was associated with smaller acceptable deviation; some reported sharp turns were associated with larger deviations. A few reported no change. This question corresponded, of course, to a question about ERRLIM. The answers appeared to support the idea of a variable performance standard, responsive to track frequency and experience.

Subjects were asked if they varied how aggressively they reacted to excessive error--a question related to the ADJUST parameter. The majority of answers were positive, though often conflicting. For example, some subjects reported larger reactions later on, others smaller reactions. A few said no variation occured.

A general question about whether control strategy varied during tracking drew answers similar to those in response to the earlier question about performance improvement--e.g., concentration, feel for the stick, anticipation. Two subjects reported they went from using preview to not using it.

6. Summary Discussion

The first research question to which these tests were directed required consideration of the quality and uniqueness of the matches between HOPE models and human behavior. HOPE models matched human behavior uniquely and acceptably well, about 95 percent of the time, both in ½ and ¼ Hz tracking conditions. The average values of RMS difference statistics were larger in ½ Hz tracking. There existed some indication of a trend toward better matching over time in ¼ Hz tracking. The results are supportive of the idea that HOPE matches were good enough to provide a basis for control strategy inferences. They also suggest a need for improvements in HOPE and in the RMS difference measure utilized, which will be discussed in Section V.

The second of the research questions of interest related to change in control strategy with time. The weight of results is quite supportive of the idea that a task-specific control strategy developed with time, in both groups of subjects. Thus, not only did RMS error decrease significantly between the first and last trials, but also control strategy changed significantly. COT, for example, decreased for both groups.

The ABA Group, who tracked a ¼ Hz track for ten of the 15 trials, also changed to use of a smaller ERRLIM between the first and last trials. ABA group subjects apparently developed with time a control

strategy involving more rapid control movements, along with a more narrow bound for acceptable performance, than they had used in the first trial.

The BAB group, whose preponderance of experience was with ½ Hz tracking, changed not only towards use of a smaller COT, but also towards use of larger values of ADJUST than the values used at first. Thus, these subjects developed with time a control strategy involving more frequent control movements as well as greater compensation for lag, than had been used in the first trial.

Also supportive of the idea that control strategy develops with time and experience were the results which showed significant differences in mean values between the two subject groups for all three CSPs measured on the last trial, but a difference only in mean COT values on the first trial. Group differences were much more exaggerated on the last trial than on the first. These results are also supportive of the idea that control strategy develops with experience.

The third research question was directed to determination of the extent to which control strategy appears to be task-specific. The results of analyses directed to this issue suggest that control strategy does indeed become task-specific. There were significant differences between the two subject groups--one measured in ½ Hz tracking and one in ¼ Hz tracking--on all three CSPs by the last trial. These differences were in directions consistent with task characteristics as well. Subjects in the more difficult ½ Hz condition were estimated to move more frequently, to utilize a broader band of acceptable performance, and to compensate more actively for the lag characteristics of the external plant than were the ¼ Hz subjects. Such differences in control strategy are quite appropriate for effective performance involving control of an external plant with a lag which increased as the control stick moved away from center, in response to higher compared to lower track frequencies.

The fourth research question was directed to discovering the extent of predictive validity that could be demonstrated with the use of control strategy measures. Results from a variety of analyses supported the idea that values of the COT and ERRLIM parameters, measured early in learning, may be used to predict values of the same CSPs later in learning. These same CSPs are also predictive, when measured in ½ Hz conditions, of their values after transfer to ¼ Hz conditions.

An effort was made to predict RMS error from measures of control strategy taken in the same or in an earlier trial. Results obtained suggested that measures of all three CSPs made in ½ Hz conditions (for either group) permitted predictions of RMS error in the same, or in a succeeding trial.

There were some puzzling aspects to the results obtained. One is that measures of CSPs made in ¼ Hz tracking did not show significant linear relationships either with the same CSPs measured in a subsequent ½ Hz trial, or with error measured in the same or succeeding trial. That is, better predictive validity was demonstrated for measures made in ½ Hz conditions than for measures made in ¼ Hz conditions. This

is surprising because the quality of matching between best-match model and human is better in ¼ Hz than in ½ Hz conditions. The explanation for this result most probably lies in the RMS difference statistic itself. As will be discussed in Section V-C.1, this statistic is limited in its power to reflect fine differences among behaviors. Subjects are more similar to each other in ¼ Hz than in ½ Hz conditions (see Section V-B.1). Therefore, the impact of the weakness in RMS with respect to discrimination of fine behavior differences may be greater in ¼ Hz conditions, and this weakness may reduce the predictive validity of those measures.

A second puzzling result is the fairly strong negative relationship between ERRLIM measured in ½ Hz tracking and RMS error. The meaning of that parameter is that of the outer bound for acceptable performance. Defined in that way, then, the expected sign of the correlation with error in any training condition is positive. That is, the larger the ERRLIM, the larger the error. The reverse relationship was observed. That is, in ½ Hz conditions, smaller ERRLIM values were associated with larger error. This result, along with the rather large values of ERRLIM measured (see Section V-D) casts some doubt on the faithfulness of this parameter's representation of human standards for performance. These doubts argue for careful analysis of the processes in HOPE which utilize ERRLIM, as will be discussed in Section V.

SECTION V

RESEARCH ASSESSMENT AND RECOMMENDATIONS

A. Introduction

One purpose of this chapter is to present an evaluative overview of the scientific issues which have been encountered in this research. A second intent is to describe, based on this evaluative overview, further research which could build on the base established in the three years' work described in this report and in Engler et al. (1980).

The scientific issues of primary concern are the following:

- the validity of the simulation HOPE,
- the quality of the current control strategy measurement procedures, and
- the validity and utility of the modeling approach to measurement of control strategy.

Following discussions of each of these, the priorities for recommended related research are presented.

B. Assessment of Human Operator Performance Emulator: HOPE

1. Ability of HOPE to Emulate Human Behavior

The similarity of HOPE model behaviors to human behaviors is an important issue in this research assessment for the following reasons. HOPE was developed to emulate one aspect of human control behavior--control input (control stick positions) over time. As a result of that emulation, HOPE also emulates system output (cursor positions) over time. The similarity between human control stick positions, recorded every 40 msec, and those produced by a HOPE model, is the basis for inferences about control strategy in humans. Control output similarities, measured as the differences between human and best-fit model cursor wave forms, represent the similarities in system performance between the two. Finally, the emulation of human behavior is important evidence for the validity and potential utility of the simulation. Therefore, the similarity of HOPE model behaviors to human behaviors is a vitally important issue.

Two measures of similarity have been utilized in the research. The first of these is average root mean square (RMS) difference between discretized wave forms generated by HOPE and used by humans. Examples of these are the control inputs and system outputs recorded every 40 msec from HOPE and from humans. The second measure of similarity utilized is visual examination of pairs of wave forms.

a. Similarities as measured by RMS difference--Three waveform pairings have been compared using RMS difference. First, difference calculations have operated on pairings of human and HOPE model control

input (control stick positions over time) in order to find the best-matching HOPE model and thus identify, or infer, a human control strategy for the given time interval. Second, RMS difference calculations have also been performed on pairs of human control input wave forms, in order to compare the degree of similarity attained by best-match models of human behavior to the degree of similarity of human behavior in the same task. Third, the similarities of both model-human and human-human pairings of system outputs (cursor positions over time) have been calculated using RMS differences. Mean RMS values for these three pairings are shown in Table 14 to 16.

TABLE 14

MEAN RMS DIFFERENCES BETWEEN CONTROL INPUTS OF BEST-FIT MODELS AND HUMANS

Group	Trial Sets Track Type	Trials 1	2	3	4	5	% Time Bins Meeting Criteria
ABA	¼ Hz	3.3	3.1	2.9	2.8	2.7	82
	½ Hz	6.1	6.4	6.3	6.3	6.4	85
	¼ Hz	3.3	2.9	2.8	2.7	2.6	83
BAB	½ Hz	6.6	6.5	6.6	6.4	6.4	86
	¼ Hz	3.5	3.2	2.9	3.0	3.1	84
	½ Hz	6.1	6.2	6.2	6.2	6.2	86

Table 14 illustrates several important aspects of the control input comparisons between model and human. A most striking aspect of Table 14, with respect to the similarity issue, is the relatively poorer matching obtained when HOPE and humans were required to track the higher frequency ½ Hz track. Even this matching is, nevertheless, quite adequate. The larger RMS difference scores for the ½ Hz track conditions represents an average difference of about 5 percent (of possible differences) compared to a 2.5 percent difference when humans and HOPE models were given ¼ Hz cutoff frequency tracks to follow. These are relatively small percentage differences; their small size suggests that HOPE models are reasonably similar to human beings in their control input (control stick) behavior.

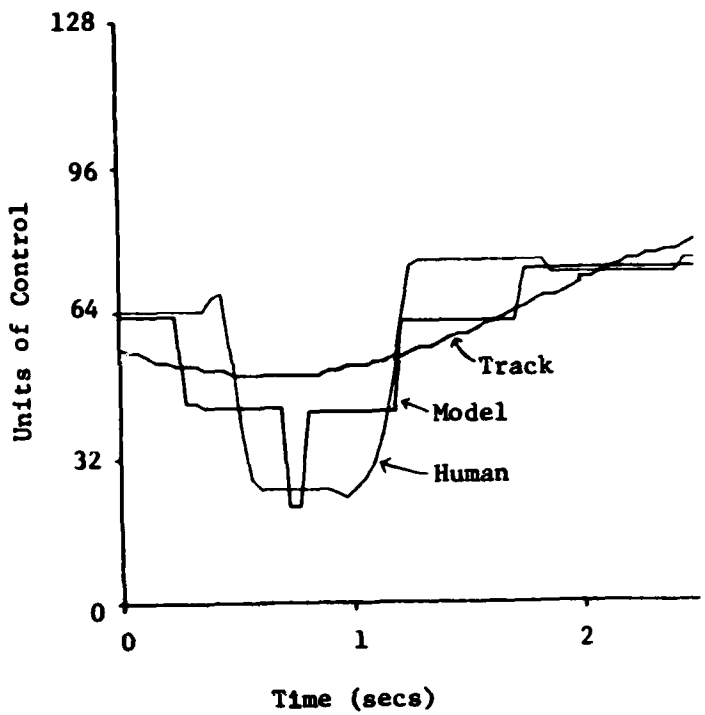
Table 15 illustrates the degree of similarity between human cursor positions (system output) and those of the best-fit HOPE models. These data reveal that system outputs from humans and their best-match HOPE model are even more similar to each other than are their control inputs. The plant dynamics which intervened between control inputs and system outputs included a variable lag which served to dampen the frequency differences observable in control input, especially in ½ Hz tracking. Indeed, differences in system outputs are hardly elevated at all in ½ Hz, compared to ¼ Hz tracking, even though the similarity between model-human control behavior was distinctly less in ½ Hz track conditions.

TABLE 15

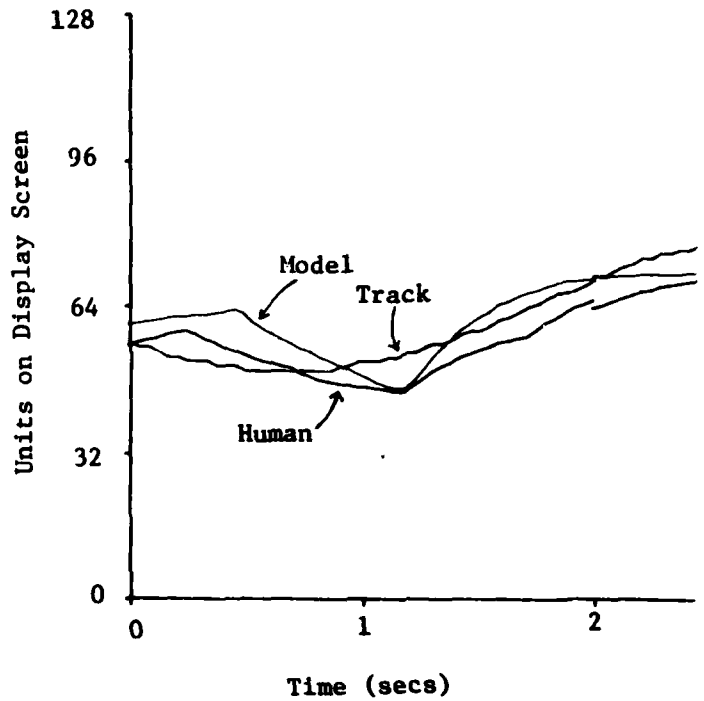
MEAN RMS DIFFERENCES IN SYSTEM OUTPUTS OF BEST-FIT MODELS AND HUMANS IN GROUP ABA

Trial Set Track Frequency	Trials 1	2	3	4	5
¼ Hz	3.0	2.5	3.3	3.2	3.0
½ Hz	3.9	3.7	3.9	3.9	4.0
¼ Hz	2.9	3.0	3.2	2.6	2.9

Table 16 provides one basis for evaluation of the model matching data shown in Figures 19 and 20. To prepare these data, RMS differences were calculated between humans from Group ABA in the validation tests. Only one group was used since responses to both track frequencies were represented in that group's behavior. These differences were calculated every 2 seconds in order to choose, for every such time bin, the human control input wave form most like the subject of comparison. The values of the RMS difference statistic were then averaged within trials and within the 12 subjects in Group ABA.

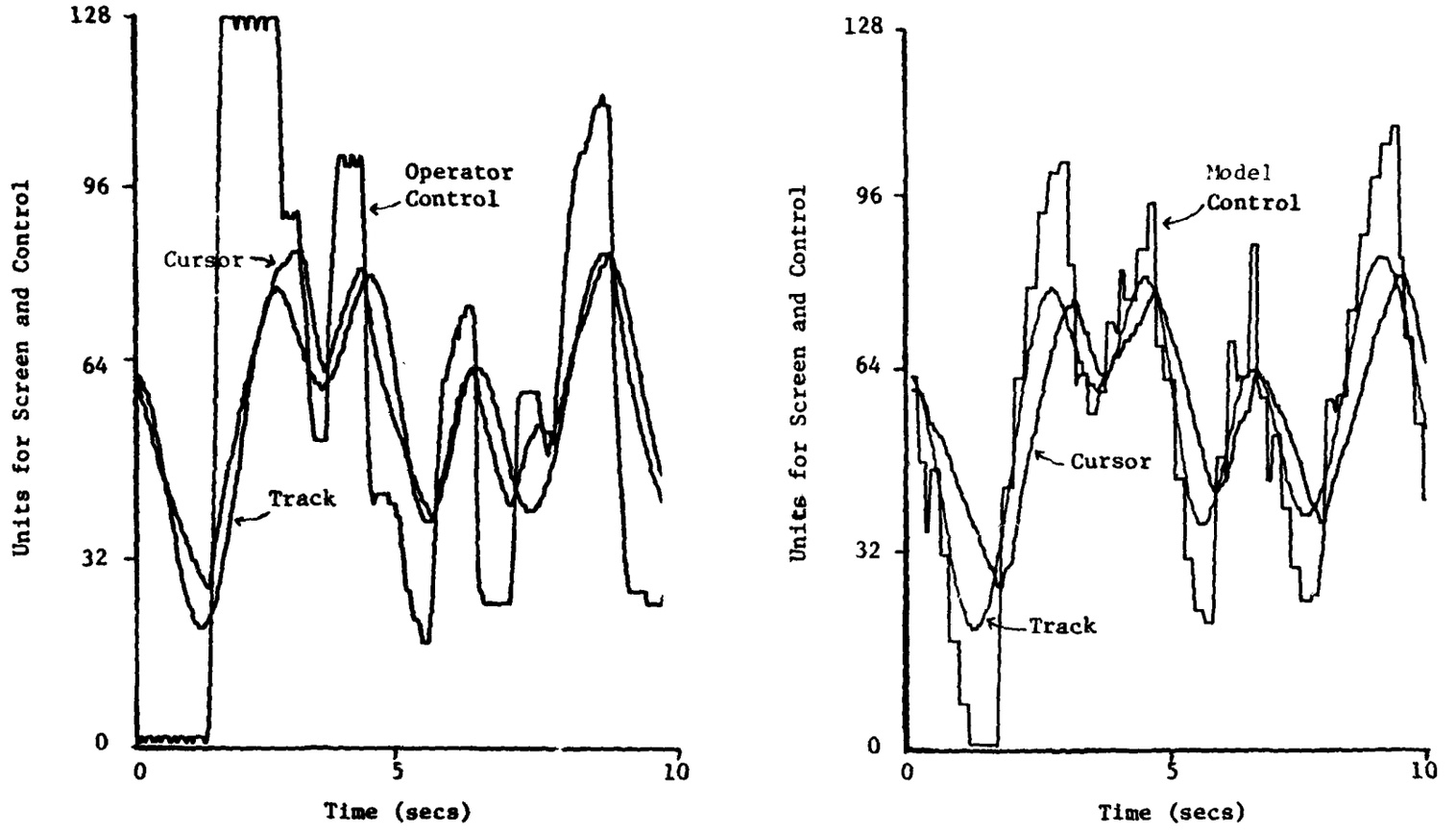


a. Human and Best-Match Model Control Stick Positions, Track Positions: Mean RMS Difference in Control Stick Positions = 12.7



b. Human and Best-Match Model Cursor Positions Resulting from Poorest Acceptable Match

Figure 19. Example of Poorest Quality Match Used for Inference



a. Typical Human Control and System Output, 1/2 Hz Track

b. Typical Model Control and System Output for Same Points, 1/2 Hz Track

Figure 20. Visible Differences in Human and Typical Model Control Style in Periods of Sustained High Rates of Change

TABLE 16

MEAN RMS DIFFERENCES IN CONTROL INPUTS OF BEST AND WORST FIT HUMANS IN GROUP ABA

Matching	Trial Set Track Type	Trials 1	2	3	4	5	% Time Bins Meeting Criteria
Best	¼ Hz	3.9	3.4	3.2	3.2	2.8	91.8
	½ Hz	6.8	6.5	7.0	6.7	7.9	90.7
	¼ Hz	3.3	3.3	3.5	2.8	2.8	87.8
Worst	¼ Hz	10.0	8.3	7.8	7.6	7.0	Not Applicable
	½ Hz	16.7	16.2	16.0	15.1	17.3	
	¼ Hz	7.4	7.4	7.2	6.0	6.3	

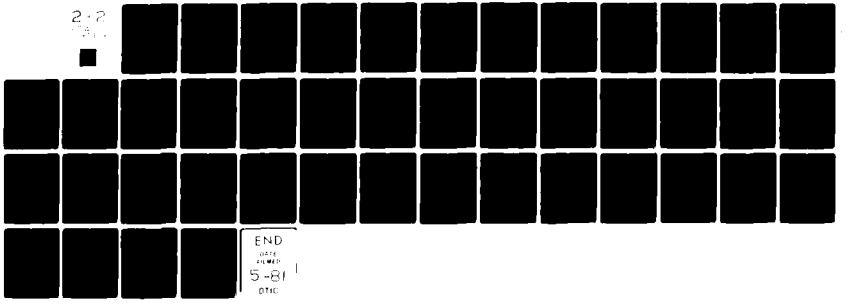
In selecting the closest human match the same quality and uniqueness criteria for inclusion in the average were applied as those used in the human-model matching process. That is, RMS difference values from human matches greater than 12.8 in any 2 second time bin, or from non-unique matches, were not included in the mean RMS difference calculation. This procedure has the drawback of artificially lowering the overall average, but has the advantage of permitting comparisons of matching qualities under identical procedures. Thus, the percentages for the closest matches indicate the percent of time bins which met both criteria, just as they do in Table 14, and the means in the two tables were calculated with the same quality ceiling.

The human-human comparisons were calculated in order to determine whether, as a class, HOPE model-human differences were grossly different in size from the differences found among human beings doing the same task. Table 16 displays these differences among humans, averaged for the best matches for each person (for those bins which met the quality and uniqueness criteria) and also averaged for the poorest human matches in those same bins.

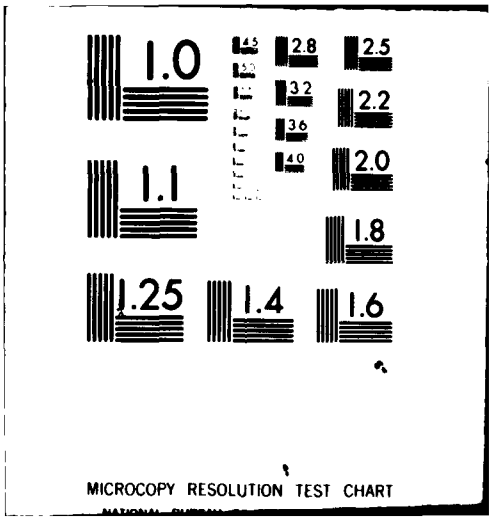
Three points are illustrated by Table 16. First, it is gratifying to note that the size of the average difference between human and human, and the percent of bins which could be uniquely matched within a 10 percent criterion were not too different from the human-model differences

AD-A097 449 GEORGIA INST OF TECH ATLANTA ENGINEERING EXPERIMENT --ETC F/G 5/9
REFINEMENTS AND VALIDATION TESTING OF HUMAN OPERATOR PERFORMANC--ETC(U)
MAR 81 E L DAVENPORT, J GREEN, W E SEARS F33615-77-C-0042

UNCLASSIFIED AFHRL-TR-81-5 NL



END
DATE FILMED
5-81
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART

shown in Table 14. Second, the average size of these differences is roughly doubled in the move from the ¼ Hz cutoff frequency track to the more rapidly varying ½ Hz cutoff frequency track. This is a pattern distinctly seen in Table 14, as well. Third, there appears to be a tendency for humans to become, on the average, slightly more similar to each other with time. There is also a comparable reduction in the best-match model-human differences, especially for Group ABA in ¼ Hz tracking conditions. The data in Table 16 support the idea that, at least using RMS difference as the measure, HOPE models can match human behavior about as well as other humans in the same training condition do.

Examination of Tables 14, 15, and 16 provides support for the proposition that the HOPE simulation emulates human control inputs and outputs rather well. This conclusion is, however, based on a single statistical measure of similarity--the RMS difference. Another measure of similarity utilized in this research is visual examination of the wave forms themselves-- control stick and cursor positions over time.

b. Visually-observed similarities and differences--Visual examination of pairs of wave forms is advantageous in that human perception can utilize more dimensions of the differences between them than does the RMS difference statistic. Visual examination is, however, severely limited in the quantity of data that can be reliably compared. Use of this technique is made, therefore, only in order to suggest hypotheses for further quantitative testing.

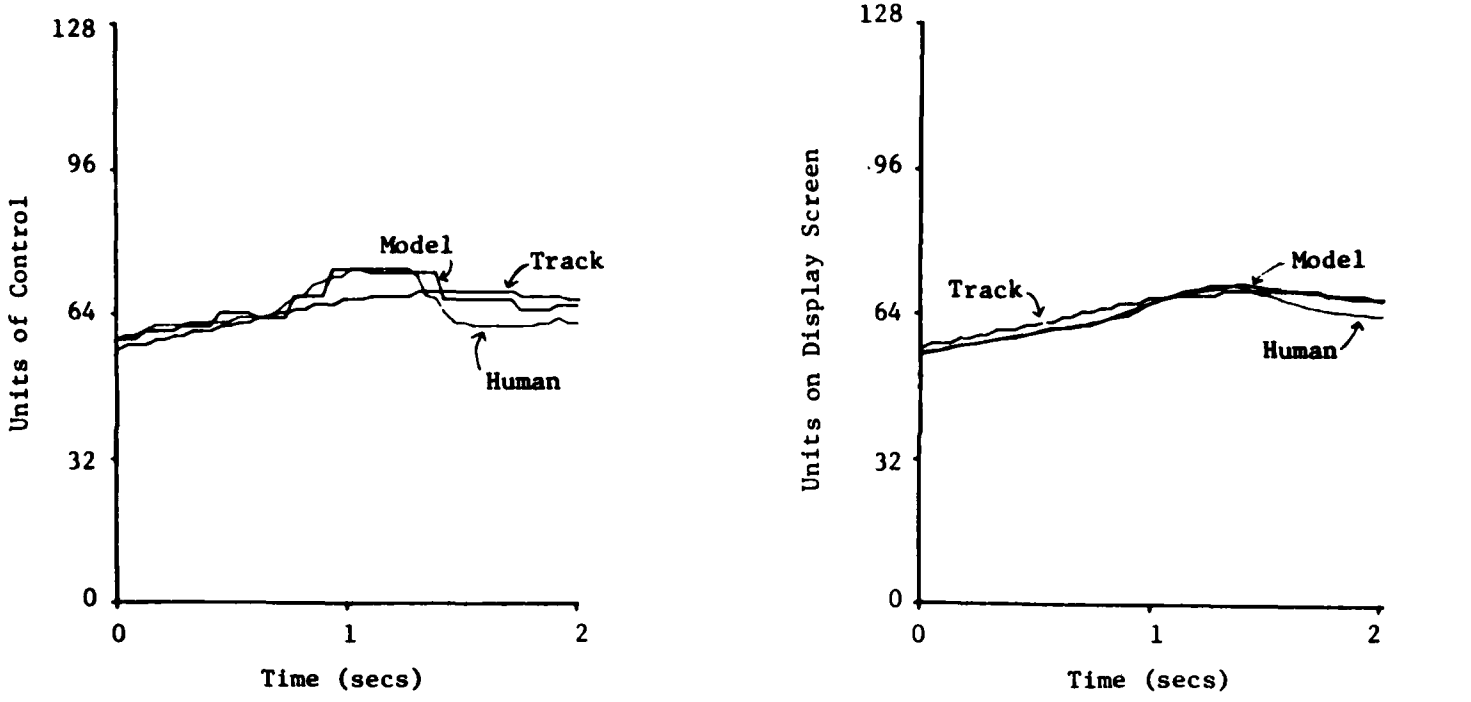
Figure 19a illustrates the visual effect of the RMS difference similarity criterion used for the control strategy measurement process. The RMS difference here was 12.7, just under the 10 percent cutoff for acceptable matching of 12.8. This is illustrative of the poorest matches used for inference. It is clear that the model and human behaved in a rather similar fashion when compared to the input wave form (the track) provided to each. Figure 19b shows the system outputs for that same time period. Similarity between outputs is much higher than between inputs. The sample chosen was selected at random and is representative of the appearance of this quality of matching. The figure fails, however, to illustrate another characteristic aspect of these data, as well. This characteristic is clearly shown in Figure 20, in which a longer period of time is shown. Similarity seemed to be lowest during periods of sustained high rates of average position change. The human tends to proceed, rather directly, from one extreme to another in such situations, seeming to resist moves in directions opposite that being taken by the track, and to resist pauses. The model, in contrast, proceeds from one extreme to another, but with several pauses and changes of direction in the process. The pauses and changes of direction inflate RMS difference most when the extremes are rather far apart, or, in other words, when the rate of change of position (velocity) is high and the high velocity is comparatively sustained. It is helpful in examining Figures 19 through 25 to remember that the RMS difference is computed in terms of the vertical distance between the wave forms, not in terms of the shortest distance.

Figures 21 and 22 are representative of the average (as indicated by RMS difference) quality of matches for the ¼ and ½ Hz track, respectively. The system output side of the figures confirms what was shown by the RMS difference measure (Tables 14 and 15); system outputs over time are more similar to each other than are control inputs. The plant dynamics which intervened betwen control inputs and system outputs included a nonlinear first order lag, which served to reduce the differences seen in control inputs. These average quality matches (in terms of RMS difference) illustrate the visual differences between model and human output shown in more exaggerated form in the poor match in Figure 19. Human control inputs are smoother and more continuous than are the inputs of the model, though both are more like each other than they are like the input wave form, also shown in Figure 19.

Figures 23 and 24 are typical of the highest quality matches achieved in ¼ and ½ Hz tracking. These figures illustrate the point that the best matching seems to take place during periods when the rate of change of position is relatively low. In these cases, style differences between human and model become much less apparent.

The organization of these samples of control inputs and system outputs has been in terms of the statistical measure of similarity discussed earlier--RMS difference. This organization was selected for two reasons. One is to provide evidence about the reasonable nature of the level of matching and the quality criterion utilized in the research. It seems intuitively clear that the RMS difference measure should correlate with the more complete visual estimation of matching quality, as it indeed does. The second reason for this order of presentation is that there appear to be no surprises in the data not revealed by RMS difference and visual examination used together. That is, RMS differences do not indicate the character of the differences between model and human, which visual examination does. But visual examination of input samples does not reveal any qualitatively superior matches not detected by the quantitative measure.

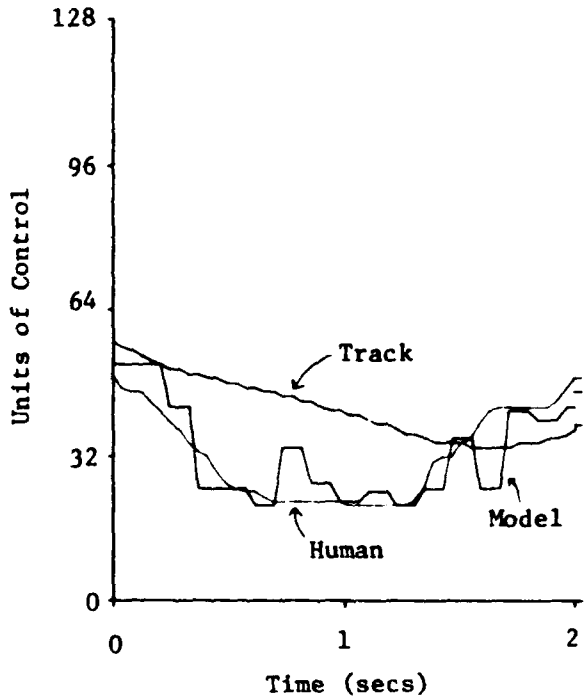
c. Summary of similarity issue--The ability of HOPE to emulate human behavior is quite high, about as high (as measured by RMS difference), as is the tendency of human beings to emulate each other when given essentially identical tasks to perform. HOPE does not, however, emulate all the qualities of human behavior--such as the tendency for smooth motion between position extremes. This latter fault is due, in part, to the absence in HOPE of any model of the physiological limits known to be present in the human neuromuscular system.
The current HOPE models only the cognitive processes associated with continuous control, and not the neuromuscular lags and frequency limitations. This absence is deliberate, since the focus of the effort at simulation development was on these cognitive processes which have not been previously explored through use of a simulation approach. Thus, the apparent severity of qualitative differences between HOPE and human control inputs is exaggerated in the data presented here. The most important issue suggested by these quality dissimilarities is the extent to which they are caused by strategy differences between human and HOPE. This issue will be explored later in the chapter in Section V-B.3, "Representation of Control Strategy in HOPE."



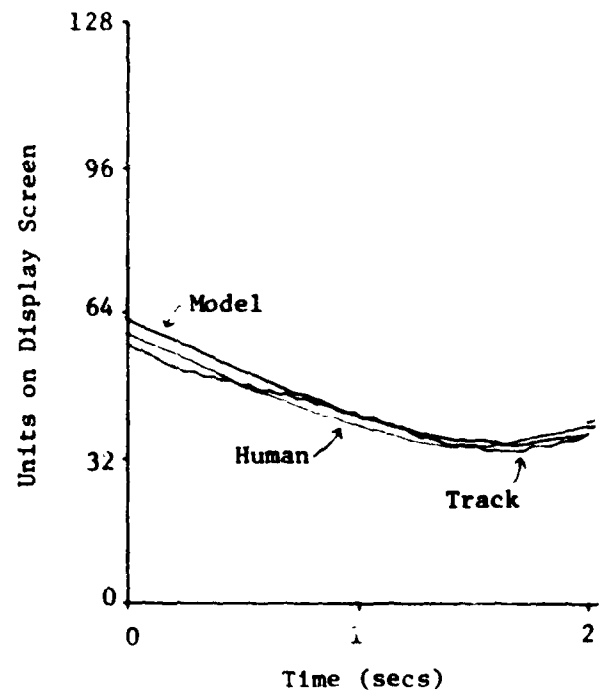
a. Human and Best-Match Model Control Stick Positions, Track Positions: Mean RMS Difference in Control Stick Positions = 3.0

b. Human and Best-Match Model Cursor Positions Resulting from Average Quality of Matching Used for Inference in 1/4 Hz Tracking

Figure 21. Example of Average Quality of Matching Used for Inference in 1/4 Hz Tracking

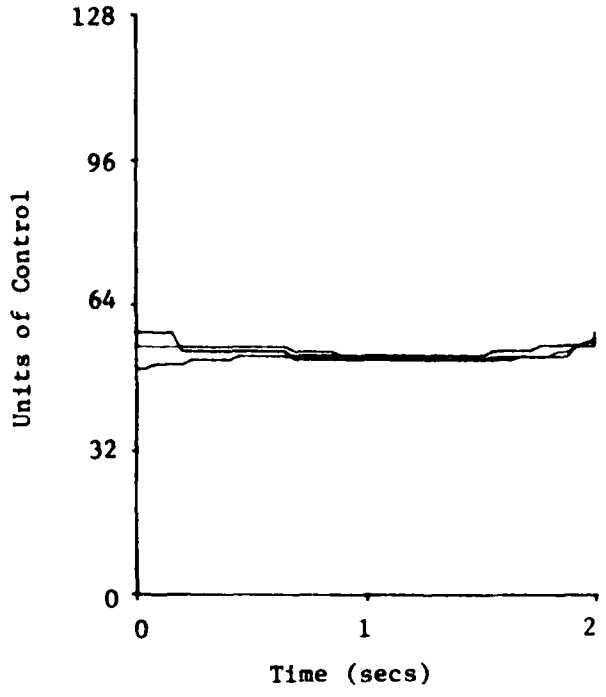


a. Human and Best-Match Model Control Stick Positions, Track Positions: Mean RMS Difference in Control Stick Positions = 6.4

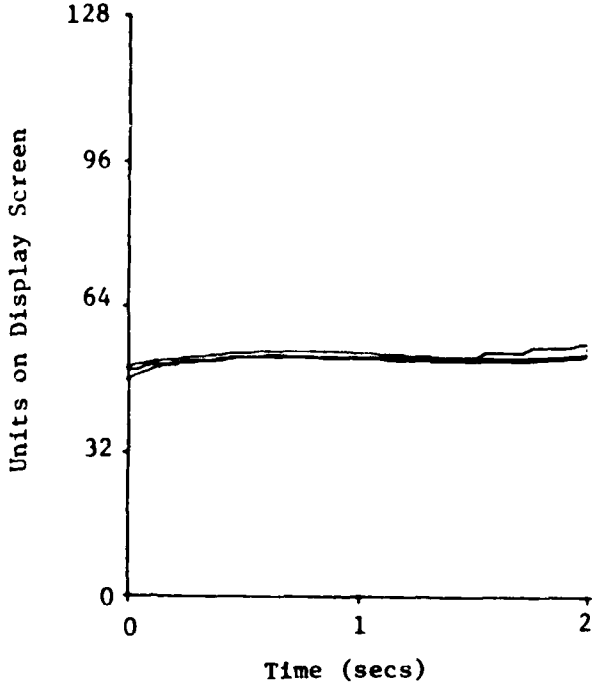


b. Human and Best-Match Model Cursor Positions Resulting from Average Quality of Matching Used for Inference in 1/2 Hz Tracking

Figure 22. Example of Average Quality of Matching Used for Inference in 1/2 Hz Tracking

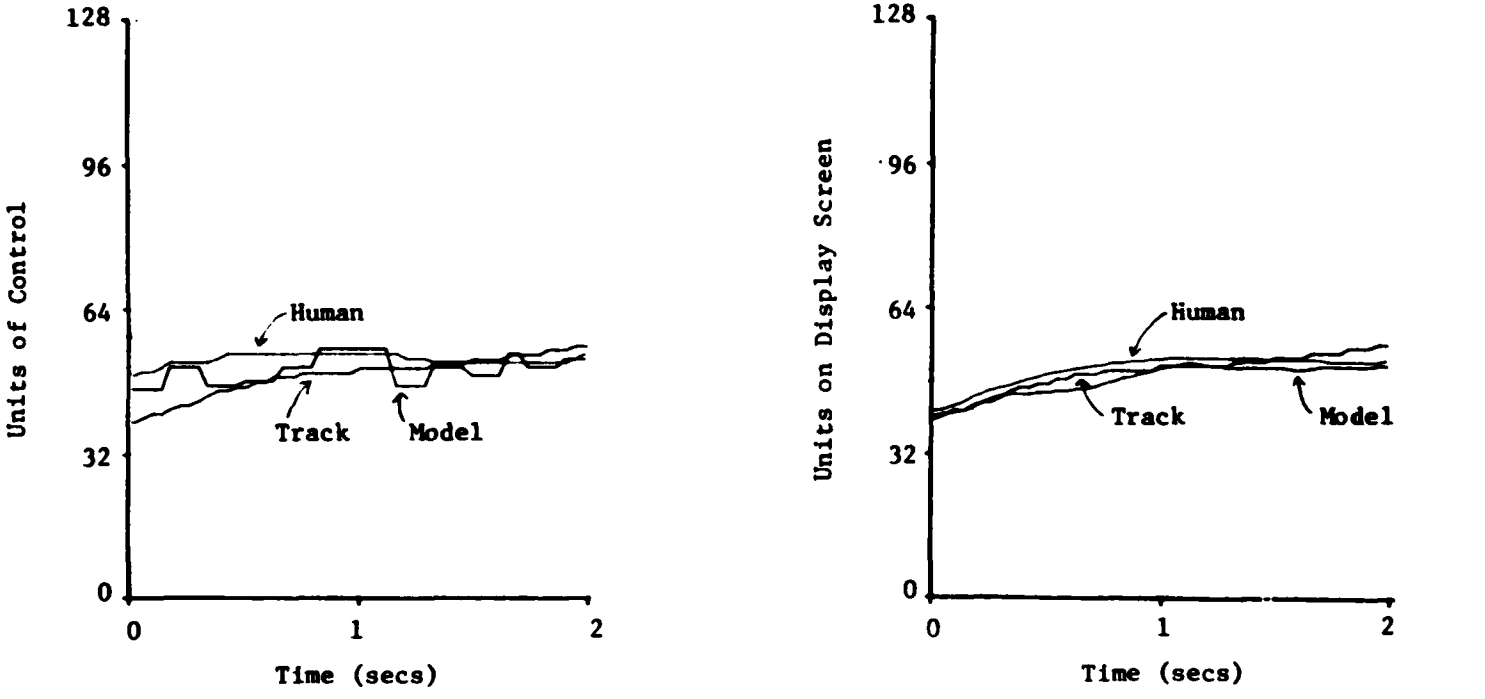


a. Human and Best-Match Model Control Stick Positions, Track Positions: Mean RMS Difference in Control Stick Positions = 1.4



b. Human and Best-Match Model Cursor Positions Resulting from Highest Quality of Matching Used for Inference in 1/4 Hz Tracking

Figure 23. Example of Highest Quality of Matching Used for Inference in 1/4 Hz Tracking



a. Human and Best-Match Model Control Stick Positions, Track Positions: Mean RMS Difference in Control Stick Positions = 3.6

b. Human and Best-Match Model Cursor Positions Resulting from Average Quality of Matching Used for Inference in 1/2 Hz Tracking

Figure 24. Example of Highest Quality of Matching Used for Inference in 1/2 Hz Tracking

2. Psychological Validity of HOPE

The Human Operator Performance Emulator (HOPE) is designed to represent the psychological organization, structures, and cognitive processes underlying human continuous control learning and behavior (see the first report on this work, (Engler et al., 1980) and Section II of this report for full details of HOPE). It is not designed to represent physiological processes, structures, and organization. There has been an explicit attempt to build in psychological validity at the global level, by including representation of such accepted psychological constructs as short and long-term memory, inference processes, processing limitations, and sources of error. There has been, as well, an attempt to represent details of these constructs in a reasonable way, with the full awareness that alternative representations of the details of these constructs are possible. Further, specification of the details of HOPE constructs has occurred with the awareness that establishing the validity of one representation of mental structures and processes over all others is not possible (see Anderson, 1978 for a full discussion of this issue). The purpose of this part of Section V is to evaluate the extent to which known invalidities or indeterminacies may affect the validity and generalizability of the control strategy measurement process central to this research.

A two part strategy has been followed. The first part involves relating previously known invalidities or omissions in the global structure of HOPE to observed data. The second means of assessing the impact of HOPE's validity on the control strategy measurement process is through examination of data from HOPE in order to generate hypotheses about previously unsuspected problems in HOPE. In other words, logic, interviews with subjects, and the psychological literature may suggest that certain problems will be evident in the data due to the incompleteness of HOPE. On the other hand, the data from HOPE, when compared with that of humans, may reveal other, unsuspected, invalidities in HOPE.

a. The relation of known incompleteness in HOPE to observed data--HOPE has no representation of the delays and frequency limits that must intervene between the intention of motor action and its actual execution. This omission was deliberate, although it may have been mistaken, and was made in order to focus on psychological rather than physiological aspects of continuous control. The interaction between known physiological limits and delay and psychological constructs such as memory may mean, however, that the omission of representation of such limits is a serious flaw in HOPE, and may directly interfere with the control strategy measurement process as it is now carried out. To see why, consider the following argument.

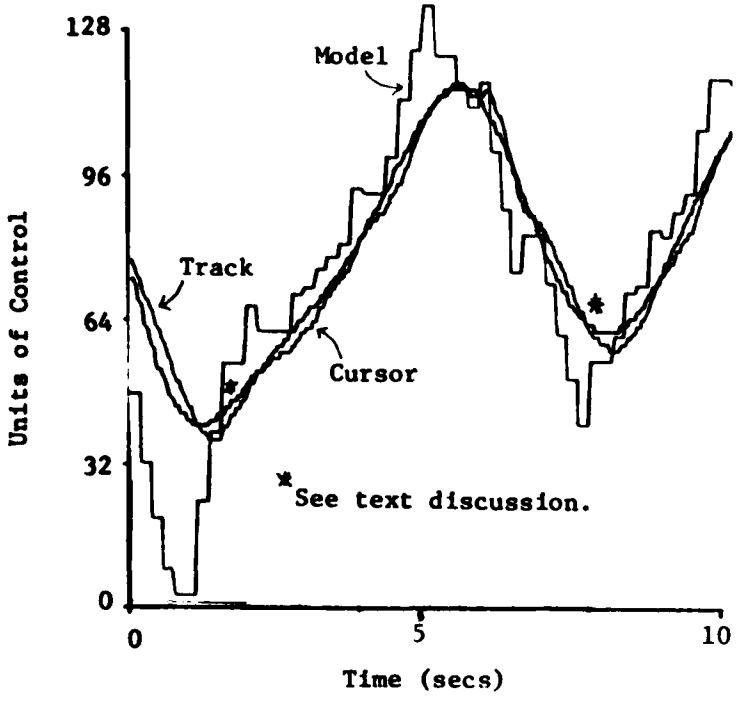
Delays and frequency limitations intervening between intentions and observable actions are likely to have at least two profound effects on entities that are very important to this research. The most direct effect is on the control input wave form (control stick position over time). Specifically, neural transmission times, inertia in muscles, and other factors will serve to eliminate nearly all of the frequency content above 7.5 Hz in human control behavior. Thus, any frequency

content of above 7.5 Hz in HOPE model control outputs should also be filtered out, for maximum similarity to human behavior. The effect of such filtering is illustrated below in Figure 25. Figure 25a shows a 10 second (simulated real time) record of control inputs from one HOPE model. In Figure 25b the same portion of a record is shown, utilizing a HOPE model in which a low pass filter has been applied to all commands issued since the model began the task. In other words, this HOPE model was equipped with a representation of neuromuscular frequency limitations intervening between its representation of intention and its representation of motor action. It is easy to see that the presence of a filter eliminates many of the characteristics of the data which were most troubling--the excess variability and "spikiness" revealed in visual examination of many of the models. It is also obvious, upon careful comparison of Figure 25a and 25b, that other aspects of HOPE have been affected by the presence of the filter.

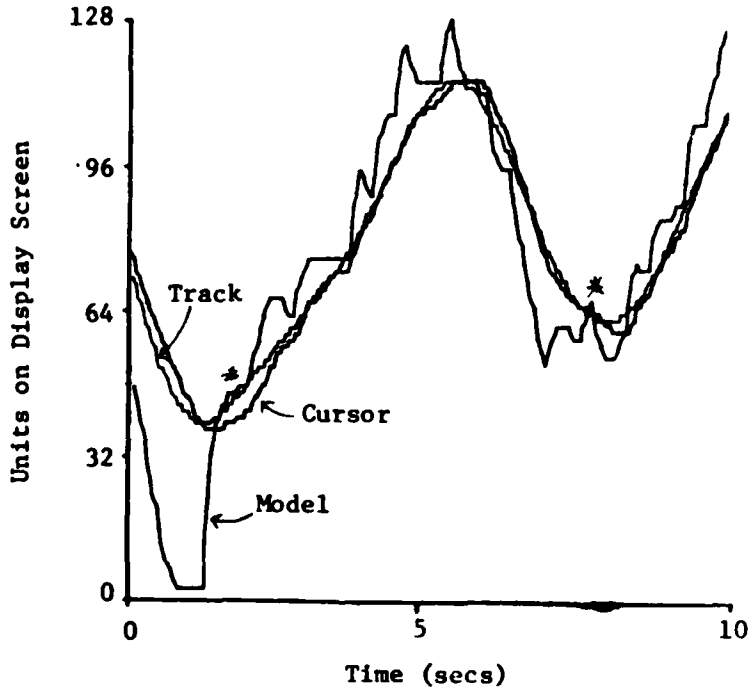
Note, for instance, the asterisks shown at identical points in simulated time in the two parts of the figure. The control stick positions at the asterisks in Figure 25b do not represent simply the effect of a filter on the positions in Figure 25a. Apparently, somewhat different intentions (commands for control stick position) were generated in response to exactly the same representation of environmental stimuli. There are several such command differences visible throughout the 10-second period, identifiable in relation to the track, which is identical in both halves of Figure 25.

Such a difference in intention must result from differences in the source of the intention in the model--the Command Memory. Directly or indirectly, the Command Memory is the source of the great majority of all commands issued. For example, on the average, 90 percent of commands used by HOPE after four trials of experience come directly from the memory. This percentage is much increased late in learning, and both are further increased when commands coming indirectly from the memory through "generalization" processes are considered. Thus, the presence of neuromuscular lags results not only in a reduced frequency content, but more than likely also in slightly different contents in HOPE memory.

Since it is evident that humans do have such those physiological limitations and that HOPE at present does not, and since our reasoning and the data suggest that the presence of such representation would result in more human-like behavior on the part of HOPE, the conclusion is that the omission of that representation may have a major impact on the validity of the control strategy measurement process--a process based on similarities in behavior between human and model. The remedy for the omission involves including in HOPE not only a representation of the physiological delays present but also a representation of human knowledge of the characteristics of these delays--an internal plant model, in the terms of the theory presented in Section III of the preceding report (Engler et al., 1980). The interaction of that knowledge with the knowledge of the vehicle being controlled--Command Memory in HOPE--is a very complex issue, and its investigation was not undertaken in this research program.



a. Control Stick Inputs, Cursor Outputs, Track of Typical HOPE Model Without Filter



b. Control Stick Inputs, Cursor Outputs, at Same Track Points of the HOPE Model in a), With Filter

Figure 25. Effects of Applying One Pole Low Pass Filter to a Typical HOPE Model as Representation of Physiological Delays and Limitations

A second deliberate omission from HOPE is that of what was termed, in the theoretical discussion (Section III, Engler et al., 1980), an input predictor. The input predictor is a representation in HOPE of human knowledge about the future sequence of desired states. In pursuit tracking of sinusoidal wave forms, for example, human subjects are soon able to anticipate the upcoming track and thus improve their performance. The task utilized in this research is, however, preview tracking rather than pursuit, and so a representation of that anticipatory ability was not included in HOPE. Subjects in their post-experiment interviews frequently mentioned that they believed that learning the track pattern helped them improve performance. Such learning is, on its face, correspondent to the learning of an input predictor. What is more likely involved, however, is some type of grouping together of response sequences in memory. Such a grouping could be represented in HOPE by the filling in of all cells in the Command Memory which correspond to a given sequence of moves, thus permitting smoother, less attention-demanding performance throughout those sequences. For this reason, it is believed the omission of an input predictor has no ill effect on the control strategy measurement process.

A third omission from HOPE was that of certain aspects of control strategy. Interviews with subjects engaged in the tracking task utilized have suggested the variation by humans of certain aspects of control strategy which are, at present, fixed in the simulation. Such omissions may well have impacted the measurement process unfavorably. This topic will be discussed in Section V-B.3.

Note should be made here, as well, that alternative representations of the Command Memory, the Excessive Error Process, and of the sources of error in memory are possible. Indeed, it may be impossible to truly identify these structures and processes through modeling. What can be accomplished is to determine those structures and processes which result in most human-like behavior, and utilize these to provide insights into the effects on behavior of various training environments.

b. Possible invalidities in HOPE structures and processes revealed by data patterns--The previous section described analyses undertaken to discover the effect of invalidities and omissions in HOPE that were known prior to testing. This section describes certain data patterns which suggest additional problems existing in HOPE--problems which should be remedied for HOPE and this approach to understanding control strategy to reach their full potential.

There are two primary problems revealed in the data.

(1) HOPE's much poorer matching quality on curving track sections and in ½ Hz tracks, including the visual distinctions between human control behavior and HOPE model control behavior--i.e., HOPE progresses from peak to peak with many reversals and "spikes" while humans, in general, do not.

(2) The differences in patterns of system performance (RMS cursor error) between humans and HOPE models, including differences between humans and their best-fit model and between humans and the models as a group.

The first problem has been discussed before, and will be discussed again in the section on control strategy omissions. Those differences in control stick behavior on track curves which remain after representation of human delays and limitations are applied are thought to be primarily the result of the omission in HOPE of certain aspects of control strategy for this task. The second data-revealed problem is discussed below.

In Table 17 the reader can see three data sets. The top line represents mean RMS cursor error in Subject Group ABA, for the three trial sets. The second line represents mean RMS cursor error for the best-fitting models to these subjects for each of the three trial sets. The third row represents the RMS cursor error for all 216 models, averaged together.

The human RMS position error shows the pattern considered before--the reduction in error over time in ¼ Hz tracking, as well as an elevation of error during ½ Hz tracking. The errors of best-fit models show no such trends. The RMS error remains at a fairly constant level for all 15 trials, showing very little elevation during ½ Hz tracking.

TABLE 17

MEAN ERROR (CURSOR-TRACK) FOR ABA GROUP SUBJECTS, BEST-FIT MODELS, AND ALL HOPE MODELS

	Trial Set 1(¼ Hz)	Trial Set 2(½ Hz)	Trial Set 3(¼ Hz)
Human Error	2.09	3.18	1.57
Best-Fit Model Error	3.64	3.96	3.64
All HOPE Models	2.31	4.39	2.40

The models as a group do show the elevation in ½ Hz tracking, but fail to show any reduction for the third trial set, as compared with the first. These data patterns may stem from a variety of causes. For example, the fact that best-match models show a comparable size error to that of humans in ½ Hz, but not in ¼ Hz tracking, may relate to what might be termed poor precision on the part of RMS difference. This problem is inherent in the RMS measure, and is discussed in Section V-C.1.a. Some evidence for this poor precision is found in the fact

that the control strategy measurements made in ½ Hz tracking show considerable predictive validity (see Section IV) while those made in ¼ Hz tracking do not. This problem might result from the crudeness of the similarity measure used, relative to rather fine distinctions necessary to differentiate human control behaviors in the simple ¼ Hz task. In other words, the ¼ Hz track moves rather slowly and smoothly, with no unexpected or difficult-to-make responses required. There was no concurrent task for these subjects, and so there were not large differences between subjects in how they responded. (As evidence for this, consider the fact that when humans were compared to humans, using RMS difference, the least similar pairs were within the criterion of 12.8 set for acceptable matching, for all 2 sec time bins in ¼ Hz tracking (see Table 16). Such was not the case in ½ Hz conditions.) Since RMS difference is sensitive only to position differences, it may not be adequate to distinguish responding in such easy conditions. Such relatively poor precision in measurement might result in the selection of models that were adequately similar by RMS difference statistics but were not sufficiently similar to mimic performance, resulting in larger error on the part of best-fit models in ¼ Hz conditions.

The other obvious problem with these records is the failure to observe a reduction in error for the last trial set on the part of HOPE models. In a theoretical sense, this observation is not a problem, since we have assumed that an important part of skill development is the development of task-specific control strategy. The models are handicapped, in the sense that they must use a fixed control strategy throughout the task while the evidence gathered in this project indicates that humans change their strategies fairly frequently.

The restriction to use of fixed-strategy models to infer control strategy may pose a problem for the measurement procedures used in this research, for the following reason. A fundamental assumption made in the procedures is that humans and models have the same amount of knowledge regarding the external plant, at any point in time at which control strategy is measured. Both humans and HOPE models begin tracking without detailed knowledge of the response characteristics of the external plant. In HOPE, that knowledge is represented by the Command Memory, a two-dimensional array in which two cursor states are related by the control stick position necessary to achieve that transition within one COT. If people frequently vary their COT, as they appear to do, then the contents in the Command Memory of a model representing the human must also change to represent the new COT. In the present procedure, a change in estimated COT is the result of a model with a changed COT being selected as best match. What may pose a problem is the fact that if changes in COT occur frequently, then the models (with fixed COTs) have denser command memories (since they have been gaining information nearly every COT since tracking began) than do humans, using any given COT for only a part of tracking. The difference in knowledge is contradictory to one of the fundamental assumptions made in the procedure. It is difficult to say how much of an impact the possible differences in human and model knowledge have on the procedures and inferences made here. It is, however, a factor that may become increasingly important as the tasks modeled increase in complexity.

3. Representation of Control Strategy in HOPE

a. Omissions--The definition of control strategy, as the parameter set determining input stimuli, output criteria, and sequencing among decision processes, is quite a comprehensive one. However, in the present HOPE not all of such parameters appropriate to the tracking task used in these studies were permitted to vary. Subjects, in post-test interviews, reported that they varied the amount of preview used, alternately attending to points several cm ahead and attending only to the point where the cursor was (see Section IV-C.5). The models do this as well, depending on how long a command string has built up in the Command Buffer, but this length is not parametrically controlled in the model. In HOPE models, command string length is a function of the value of ERRLIM and the current position of the cursor in relation to ERRLIM. Human subjects, however, reported use of increasing preview as a function of track frequency; e.g., the higher the frequency, the longer the preview. HOPE did not increase its command string length in this way. Subjects also reported they strove for smoothness in movement of the control handle. No such criterion restricted HOPE's Command Generation Process.

Subjects also reported varying the point on the track they were aiming for--sometimes points aimed for were quite close together, sometimes very far apart. Aimed-for points seem to be different from size of preview since the term seems to be related to the distance to be covered in a single step. In terms of HOPE, aimed-for points seem equivalent to points one COT apart. If so, the range of COT values used in HOPE may have been too narrow. The points at which the models aimed were determined strictly by COT, and the farthest point at which a model could aim is a point 240 msec away from his present (or predicted) position. Subjects seemed to aim for peaks in the track. Yet, ½ Hz tracks might achieve excursions from one extreme to another no more often than once every second. None of the COTs used in current modeling could represent aiming for such an extreme point. This "aimed-for point" may also be related to the subject-mentioned variation in preview used, since an increase in the size of a single step could also result in use of more of the available preview. If so, an increase in COT, which is sensitive to frequency, may serve both to permit not only peak-to-peak tracking on occasion, but also as greatly increased preview. There is a clear need to resolve this issue in further research.

Furthermore, not all subjects reported learning the control dynamics. It may be that some subjects only attend to learning these relationships after error has begun to decrease as a result of subject changes in control strategy. HOPE models begin to fill the Command Memory immediately, and do so as close to every COT as performance permits. The frequency of making associations in memory may also be an aspect of control strategy, a parameter, which should be permitted to vary in HOPE.

Finally, a quite important aspect of control strategy has been deliberately omitted from HOPE. That is the control strategy development process itself. HOPE models come to the task with no knowledge of the control dynamics of the external plant. They gain that knowledge as they track, using a fixed control strategy provided at the outset.

Humans, in contrast, develop their knowledge of the external plant dynamics using a variety of control strategies. Their knowledge may be dependent on their control strategy history. To the extent that the developmental course of human control strategy, as distinct from fixed values of control strategy, affects the knowledge of the external plant control dynamics, the validity of the measurement approach tested here is threatened.

b. Validity of the representation of control strategy in HOPE--The validation tests described in Section IV of this report were primarily aimed at consideration of the validity of the control strategy representation present in HOPE--Command Operative Time (COT), ERRLIM, and ADJUST. The evidence discussed in detail there showed that control strategy differed strongly between subjects in different groups at the end, but only slightly at the beginning of training. In particular, by the end of training, COT was shorter in the faster changing ½ Hz track, ERRLIM was larger, and so was ADJUST. These differences are good support for the HOPE representation of these parameters, since the characteristics of the two tracks suggested that differences should be in those directions. Furthermore, measures of control strategy made in the fifth trial of tracking were significantly predictive of measures taken in the 15th trial. Measures taken in ½ Hz tracking, but not in ¼ Hz, were significantly predictive of measures of control strategy made after transfer to an alternate track frequency. Also, measures of control strategy made in ½ Hz tracking, after at least three trials, were predictive of both same and following trial error. In particular, COT was related positively to error; ERRLIM negatively. These findings are largely supportive of the validity of control strategy representation in HOPE.

Findings which are less supportive of the validity of current representations include the absence of predictive validity for measures of control strategy taken in ¼ Hz tracking, and an unexpected inverse relationship between estimated ERRLIM and error in ½ Hz tracking. A second set of negative findings relates to the range of CSP values selected. While mean values for ADJUST and ERRLIM are appropriate, the range of these values suggests there may be a problem in their measurement or their representation, or both. Table 18 shows the frequency of values picked for each CSP. The range of measurement for ERRLIM is a problem because there are so many bins in which ERRLIM is too large for plausibility, in light of the guidelines presented and in light of subject reports. These latter data suggest that few, if any, of 2 sec time bins of tracking should be characterized as having ERRLIM values outside the presented guidelines (larger than 1.6 cm); yet many such bins are so characterized by the present HOPE. This result may suggest improvements needed in the way HOPE evaluates and reacts to error--the Excessive Error Process. High values of ERRLIM used in HOPE result in very infrequent operation of the Excessive Error Process, thus calling into question its validity.

The range of estimates is also a problem for the ADJUST parameter. ADJUST values are frequently estimated to lie at the low end of the range permitted, as shown in Table 18. While there is less evidence than for ERRLIM regarding the appropriate size for ADJUST, the small

or zero measured values, combined with less consistently meaningful variation than in COT and ERRLIM, suggest that this parameter representing compensation for lag, or reaction to excessive error, may not operate in HOPE quite as it should.

TABLE 18

PERCENTAGE OF BEST-FIT MODELS IN TRIAL 15
USING EACH VALUE OF CONTROL STRATEGY PARAMETERS

Parameter	Value					
COT (msec)	40	80	120	160	200	240
Group ABA	2.3	18.4	26.4	14.7	24.5	13.7
Group BAB	7.8	49.9	20.7	7.2	7.2	7.2
ERRLIM (cm)	0	.59	1.18	2.36	4.72	9.44
Group ABA	17.0	13.8	14.9	17.0	16.5	2.0
Group BAB	8.5	9.3	13.1	18.9	28.2	22.1
ADJUST (control units)[a]	2	4	6	8	10	12
Group ABA	28.5	19.6	13.4	14.4	11.7	12.3
Group BAB	22.1	16.6	15.4	14.7	16.5	14.7

[a]There are 128 distinct positions for the control stick.

4. Conclusions About the Effects of HOPE Invalidities on the Control Strategy Measurement Process

The control strategy measurement process is based on an assumption that all aspects of control strategy are somehow represented in control stick output. Similarities in control stick output between humans and HOPE models are used to infer human control strategy. These ideas taken together suggest the problems in HOPE structure, processes, or behavior which most seriously impact this line of research are:

- obvious dissimilarities between human control output and HOPE model output, particularly evident in control behaviors in response to frequent curves,

- omissions of the internal plant and the internal plant model, the representations, respectively, of human physiological delays and limitations and human ability to take them into account,

- omissions of aspects of control strategy in HOPE, aspects such as amount of preview utilized, which are reported to vary among humans engaged in the experimental task being utilized,

- the relatively high percentage of estimates of implausibly high values for human ERRLIM, and implausibly low values for ADJUST, and

- the negative relationship between ERRLIM and error.

These problems should be resolved for HOPE to realize its full potential as a measurement tool. They do not, however, outweigh the amount of evidence supportive of the HOPE utility and validity. Control strategy, as measured by HOPE, showed, on the whole, meaningful variation with learning and with changes in training conditions. The changes and their frequency were consistent with subject reports about their strategies. Evidence of predictive validity was also obtained. This evidence is quite strong and positive for the early stages of such an ambitious undertaking.

C. Assessment of Current Control Strategy Measurement Procedure

1. Strengths and Weaknesses of Currently Used Procedure

a. The measure of similarity--The RMS difference is used to determine similarities of human and model control stick output. It is the best measure that has been found for this purpose. For example, low RMS scores are good indicators of similarity as determined visually. High RMS scores are valid indicators of poor matching as determined visually. There appear to be only insignificant differences in the average values of CSPs inferred using RMS and those inferred using other, more complicated statistics. The measure has, however, several weaknesses which threaten the validity of control strategy measurement based on this measure.

Perhaps the most serious is that while RMS differentiates easily between very good and very bad fits, it does not distinguish finely among fits in a consistent way. Data leading to this conclusion are shown in Table 19, where the linear correlations between first and second choice CSPs are shown, for ¼ and ½ Hz tracking. It was expected that the variations across time in first choice CSP estimates should be mirrored in variations across time in second choice CSP estimates. The linear correlation is a measure of such mirroring. These low correlations may be due to the fact that the RMS difference score does not scale similarity finely enough for this research. There may be other possible causes for these low correlations, but they are strong support for the search for an improved quantitative measure of similarity.

TABLE 19

MEAN LINEAR CORRELATIONS BETWEEN FIRST AND SECOND BEST FIT MODELS' CONTROL STRATEGY PARAMETER VALUES, LAST TRIAL

	Subject Group (Last Trial Track Frequency)	
Control Strategy Parameter	ABA (¼ Hz)	BAB (½ Hz)
COT	.33	.45
ERRLIM	.20	.37
ADJUST	.21	.38

A second weakness with the RMS difference score is that about ten to fifteen percent of the time the best choice picked is not a unique one. Tables 4 and 5 show the decrease in percent of bins acceptably matched that occurs when "acceptably matched" includes a uniqueness criterion. It is believed that a measure which takes into account slope and perhaps also acceleration differences would greatly reduce such uniqueness problems.

b. Other aspects of the procedure--The two other important aspects of the measurement procedure itself are the interval over which waveform differences are matched, and the criteria used to determine when matching is good enough to accept inferences about human control strategy. Both may affect the inferences made quite profoundly.

During the preliminary testing of the approach used in this research, it was assumed that human control strategy probably only varied every 20 sec or so; and, therefore, its measurement using HOPE models might also be carried out only that frequently. Examination of human control behaviors, however, in relation to the track being followed, suggested that control strategy might change much more often than that. Indeed, when the interval was reduced from 20 to 2 sec, the average RMS score dropped dramatically (see Section III). The reduction of the matching interval not only enabled much better matching of certain parts of human behavior, but also served to emphasize the dramatic differences between HOPE models and certain other aspects of human behavior. The range of RMS scores attained was much increased, with both much lower and much higher single interval scores observed. These results suggest

that for this task, then, with no other tasks competing for the operator's attention, the 2 sec matching interval is quite appropriate.

The criteria for determining when best-match model control behaviors may be used to infer human control strategy have been made more strict as HOPE and the control strategy identification procedures have been refined. The extreme of this trend to increasing strictness would be that a single model must produce a zero value on the matching statistic used, in order for inferences to be made. HOPE is not yet perfected to an extent to permit this strict a demand. The criteria presently used, however, do result in a quality of match used for inference that is similar in value to the quality of match achieved by the two humans most similar to each other, as is shown in Figures 19 and 20 of this Section. The criteria result in rejection of choices that are grossly different from humans, as indicated by visual examination, and also result in rejection of non-unique choices. The inferences made on the basis of these criteria seem reasonable, and the criteria seem appropriate at this time.

c. Analysis of CSP values for best-fit models--The adequacy of the range of the CSP values currently utilized in HOPE was evaluated by examining the frequencies of CSP values inferred to be those of human subjects. Of particular interest was whether clusters occurred at the edge of the range of CSP values tested or if most values fell comfortably within that range. For example, do the COTs of best-fit models cluster at 240 msec, the COT upper extreme in current testing, or is the clustering at a more intermediate COT value? The former case would suggest that the range of COTs tested might need to be expanded to include larger COT values. Clustering of CSPs at intermediate values would support the adequacy of the CSP values currently being tested.

Table 18 reveals that the range for COT appears adequate, since the modal values for both subject groups are not at extremes. Subject interviews and visual examination of control stick records, however, suggest that much larger values of COT are used under some conditions (see earlier discussion in Section V-B.3).

Table 18 also shows clustering at low values for ADJUST for both groups and at high values for ERRLIM for the ½ Hz group. These results do not provide unequivocal support for extension of parameter ranges, however. ADJUST, for example, might be negative in a high gain dynamics situation, but would certainly never need to be so in controlling a lag-type plant. Even use of the zero value seems inappropriate for these lag-type dynamics. Larger values of ERRLIM, too, seem so unlikely in this situation that the clustering casts doubt more upon processes in HOPE that use ERRLIM than upon the range of parameters used.

In summary, the clustering shown suggests the possible inclusion of a zero value for ADJUST, as well as the need for careful analysis of the effects in HOPE of processes which use ERRLIM and ADJUST.

2. Impact of Procedures on Validity of Control Strategy Measurement

The most serious problem in the current procedures appears to be the absence of a completely satisfactory quantitative measure of similarity between discretized waveforms. Some preliminary investigations have been carried out (see Section III), but much more work should be done. The need is severe, because the measure is the base of the modeling approach to measurement. The measure must reflect human judgements, and it must scale similarity in some fashion.

Identification of such a measure is by no means a straightforward task. Research is needed for at least two reasons. There is evidence that human judgements of similarity are quite variable. Also, discretization of the data means that calculation of statistics representing velocity or acceleration is not straightforward.

With knowledge of the severity of the problems with the RMS difference measure on the one hand, and with knowledge of the very positive results from its use in the validation tests on the other hand, it is quite difficult to assess the impact of RMS difference score weaknesses on this total control strategy measurement approach. It is clear, at least, that its potential impact is quite high. The positive validation experiment results, however, suggest that the impact of RMS difference score weaknesses on this research has been moderate.

D. Assessment of Modeling Approach to Measuring Control Strategy

1. The Value and Applications of Measuring Control Strategy

It has elsewhere been argued (Engler et al., 1980) that measurement of control strategy, as defined here, should be very useful, both in the design of cost-effective flight training simulators and in the specification and individualization of the procedures which utilize them. This idea seems quite reasonable, for several reasons. High performance flying involves multiple, concurrent attention-demanding tasks all of which compete for the limited mental processing resources of the pilot. The pilot must learn to manage these resources in order to meet his goals.

Such management by the pilot necessarily involves the determination of an adequate level of performance in all subtasks, a specification of a visual, and possibly auditory scanning pattern, decisions about which configurations of stimuli are meaningful for each subtasks, and a choice of response style, in terms of effort and smoothness of control motions. Such management choices are, in fact, exactly what has been defined here as control strategy. It is certainly plausible that total performance is a function of the net effect of these determinations. Therefore, measurement and understanding of these determinations, these aspects of control strategy, seems *prima facie* an ultimately useful endeavor.

"Ultimately" is a key word in this conclusion, however, since the value of measurement of control strategy in an applied setting cannot be realized until three conditions are met:

- valid measurement procedures are established in a complex task environment,
- measures of control strategy are shown to be reliably and consistently related to performance, and
- factors involved in control strategy learning are fully understood.

The value of measurement of control strategy is not a difficult idea to accept. On this premise, the next issue of importance is the method of measurement. While several approaches to the measurement of control strategy might be attempted--e.g., eye movement recordings, statistical summaries of control style, or self-reports--a psychologically based modeling approach to measurement has been adopted in this research, based on its potential advantages over other methods (see Section V-D.2.b). This innovative approach has been devoted to initial aspects of these three conditions. For the full realization of utility in an applied setting, the modeling approach should be validated in more complex settings, based on its initial successes in the single task environment.

The modeling approach to measurement of control strategy requires somewhat more complex assumptions than do other approaches that might be taken. These are discussed next.

2. Feasibility of Measuring Control Strategy Using a Psychologically-Based Simulation

a. Necessary assumptions--A fundamental assumption in this research is that when individuals use distinct control behaviors, in response to the same task demands, at like stages of knowledge about the external plant control dynamics, these individuals are using distinct control strategies. This means, for example, that two trainees whose control moves are measurably different from each other must be using different control strategies. A corollary to this assumption is that if the control strategies of trainees at the same stage of knowledge about the external plant are the same, their control stick output, in response to the same task demands, will be equal. This means, for example, that the control moves of a pilot using vestibular as well as visual cues as basis for response will be effectively the same as those of another pilot using the same cues as bases for response, given the same environmental inputs to each.

A second fundamental assumption is that when control strategies are distinct, and the knowledge of plant response is equivalent, then control output will also be distinct. Thus, a pilot using only visual position information will utilize different control responses than a pilot using both visual position and vestibular acceleration cues. A corollary to this assumption is that if control outputs of individuals

with equivalent plant knowledge are not measurably distinct, then neither are their control strategies. Thus, it would, for example, be impossible for trainees with identical control response to have different strategies. It would not be possible, for example, for an increase in ADJUST to cancel out the effect of a decrease in ERRLIM so as to produce equivalent response patterns.

In summary, the two assumptions and their corollaries are (under the assumption of equal plant knowledge) listed below.

(1) If control stick motions of individual A and individual B are not equal, then neither are their control strategies.

Cor. If control strategies of A and B are equal, then so are their control stick motions.

(2) If control strategies of A and B are not equal, then neither are their control stick motions.

Cor. If control stick motions of A and B are equal, then so are their control strategies.

The measurement procedure used in this research and, in fact, of any simulation/modeling approach to control strategy measurement, has been based on assumption (1) and the corollary to assumption (2), in addition to the fundamental assumption of equal knowledge of the dynamics of the vehicle being controlled. To the extent that these assumptions are invalid, either in human beings or in HOPE (or in any other modeling approach to measurement), the value of the approach is threatened. The degree of threat depends, of course, on the extent to which the assumptions are violated. Evidence of construct and predictive validity obtained in these early tests are indirect evidence of the reasonable nature of these assumptions. It is, however, important to test the assumptions directly, where possible, including tests devoted to determining how best to measure an individual's state of knowledge about the control dynamics of the external plant. Testing the assumptions also requires the development and utilization of alternative ways to measure control strategy. These research needs will be considered later in Section V.

The ideas considered just previously suggest the importance of examination and testing of the assumptions underlying any modeling approach to measurement, or, for that matter, any research approach taken to explore previously unexplained areas of human thought and behavior. Such caution does not, however, mean that the approach taken--the use of a psychologically based simulation to teach us more about complex human information processing--is not useful. On the contrary, the psychologically based simulation has several advantages, which are discussed next.

b. Advantages of a psychologically-based simulation for control strategy measurement--A psychologically based computer simulation such as HOPE is limited in its representations only by the flexibility of computer code. The code can be written

to include representations of a wider variety of mental structures/processes than are mathematically describable. Valid representation of psychologically important processes/structures increases the utility of a model of human performance, and the likelihood that fundamental assumptions are met.

The current HOPE has the following characteristics which allow it wide and useful generalizability of application to measurement and prediction of a variety of control behaviors.

- HOPE can measure control strategies of both trained and untrained operators and contains psychologically-based representations of the differences between these. Measures of control strategy made by HOPE have shown predictive and construct validity in two independent experiments. The theory on which HOPE is based assumes that learning adds information to both system models (HOPE's external plant model) and forcing function models (HOPE's input predictor). The current HOPE models the development of knowledge about the external plant control dynamics, and can easily be expanded to model the development of knowledge about the forcing function. The theory behind HOPE also assumes that control strategy changes with learning. At present, HOPE can model behavior varying as a function of the interaction of a variety of control strategies with developing knowledge.

- HOPE has demonstrated the potential to predict the details of human behavior over time, based on its use to measure control strategy early in learning. This phenomenon means there is promise of substitution of HOPE models into various training conditions, in order to select the best conditions for individuals or groups.

- HOPE can model the development of skilled control over nonlinear external plants and plants not mathematically describable. This is possible because the Command Memory directly stores commands effective for specific state transitions. Commands are not generated through use of mathematical equations.

- HOPE models human control inputs as well as system outputs. Measurement and prediction of human control behavior provides a more precise description of behavior than does measurement of system output. The latter includes effects of system variables, as well as of human variables.

- HOPE can model a very rich representation of human control strategy. Control strategy is comprised of variables that vary between individuals and over the course of learning. The representation of control strategy in HOPE is consistent with empirical data describing manual control learning and has considerable psychological validity. HOPE's measurement procedure allows inferences about human control strategy.

HOPE has the potential to identify optimal control strategies for specific tasks, which could then be used for design of effective training devices and programs. HOPE has the potential to measure effects on behavior of different cues and cue combinations present in simulators, so that simulator design can be more soundly based.

- HOPE models preview tracking behavior, and could easily be modified to model compensatory or pursuit tracking. HOPE uses preview information to build up command strings for use in the future.

- HOPE models behavior associated with a variety of criteria. The representation of control strategy includes a criterion guiding aspects of system output--an acceptable error criterion (ERRLIM). HOPE also includes aspects of control strategy which are criteria for control input aspects of performance (e.g., frequency of control stick position shifts, COT).

- HOPE was designed to measure changes in behavior occurring with learning, and includes representations of processes/structures which are believed to change with learning.

E. Prioritization of Recommendations for Further Research

1. Introduction

The preceding discussions, in Sections IV and V, have revealed the essentially positive nature of the results obtained in this research. In addition, a variety of areas for further research have been identified, based on analyses of the scientific issues of primary concern in this research. These issues are: the validity of the simulation HOPE, the quality of the current control strategy measurement procedures, and the validity and utility of the modeling approach to measurement of control strategy. Consideration of each of these issues has resulted in recommendations for further research. Prioritization of these recommendations is made with the following criteria in mind:

Priority should be given those areas which are necessary for early demonstration of the utility of a psychologically-based simulation in simulator design or flight training setting. Those areas of research necessary include those focused most directly on increasing the likeness of HOPE control outputs to those of humans, and on increasing the ability to quantitatively distinguish among discretized waveforms. Also included are research thrusts aimed at decreasing the dissimilarities between the laboratory environment and the application environment--the flight training simulator. The research areas which best meet these needs are described next. The most needed extensions to the research reported here are those tasks mentioned in Section V-E.2. Positive results from those activities would provide the basis for prioritization among these areas listed in Section V-E.3.

2. Recommended Research Areas

a. Reduction of dissimiliarities in control stick output--Certain of the most prominent dissimilarities in control stick output between humans and HOPE models have been identified--i.e., excessive high frequency variability of HOPE models, compared to humans, especially in track segments that involve sustained high rates of change. A portion of these dissimilarities is, with high probability, due to the absence in HOPE of a representation of the internal plant and the internal plant model. The latter represents the human ability to take into account the limitations of his own body. There remains a portion of such behavior for which the explanation is not yet clear.

What is needed to remedy these dissimilarities is a research plan involving the following tasks.

- Implementation in HOPE of representations of the internal plant and the internal plant model.
- Use of current measurement procedure and already-collected preview tracking data to re-measure control strategies.
- Identification and analysis of remaining types of dissimilarities. This task would include identification of which processes and structures in best-matching HOPE models produced the dissimilar portions of tracking behavior, as well as identification of commonalities in the task setting among regions of dissimilarity.
- Modifications in HOPE to remove the causes of consistent types of dissimilarities, followed by testing on existing data. The modifications might include, for example, alterations of the existing use of ERRLIM, or even expansion of the representation of control strategy in HOPE.

b. Increased discriminating power of the quantitative similarity measure for discretized waveforms--At present, control strategy inferences are made using the RMS difference measure. Its weaknesses include an inability to finely distinguish among closely similar waveforms and an absence of representation of dimensions of similarity that are represented in human similarity judgements--i.e., velocity and acceleration. What is needed to remedy this problem is additional research including the following tasks.

- Specification of relevant dimensions of similarity. This specification should be based on human judgements of similarity between waveforms. One possible approach to this is to involve trained pilots in preview tracking, introduce disturbances to their control inputs, and determine which dimensions of disturbance are most readily detectable by these trained individuals. Other approaches, such as visual comparisons and ranking of similarities, might also be included.

- Revision of quantitative similarity measure so as to reflect the relevant dimensions. The special problems of calculations involving discretized waveforms would be addressed in this task, as well.

- Testing of revised measure with respect to its ability to replicate the data on human judgements of similarity and differences.

c. Demonstration of the effects on measured control strategy of increased task loading--The current work has, of course, been carried out in a single task setting. Applications of this measurement approach will always be in a multi-task environment. Therefore, an important research area is one described by the following tasks.

- Selection of additional tasks to be integrated with the existing preview tracking task. At least one of the additional tasks would include aspects known to be related to operator information-processing load.

- Development of experimental and data collection equipment and control programs.

- Design of experiment to study the effects of task loading on control strategy. The predictive and condition-related relationships identified in the current research would be utilized to structure this design as would the interest in task loading effects. That is, an investigation of effects of task-loading would include determination of the extent to which the relationships observed in the single task setting hold for the multiple task setting.

- Data collection and analysis. These analyses would be based on control strategy measurements made using the HOPE models and waveform similarity measure developed in research areas a and b.

3. Other Research Areas

The research and assessment described in this report suggests several other important areas for research. There are several which directly build on these results, if followed by the program described in Section V-E.2.

- The need to develop means for directly measuring the extent of an individual's knowledge of control dynamics--This activity could provide an important check on one of the basic assumptions underlying the modeling approach--that of equal plant knowledge. Also, because this knowledge appears to be so important to motor skill learning, refinement of means to determine the extent of this knowledge would be quite valuable independently of the modeling approach to measurement of control strategy.

- The need to develop and implement alternative measurements of control strategy, using other than a modeling approach, and to use these measurements to test the fundamental assumptions underlying the modeling approach--Other methods of measurement of control strategy lack the potential for development of means for prediction of the details of behavior possessed by HOPE. Their development and validation would be at least as costly as the full development and validation of the modeling approach. Their application in complex environments would be prohibitively equipment-intensive. The value of the development of other measures lies in their use in the development and validation of the causal model of control strategy--the model which could then be dynamically integrated into HOPE.

- The need to relate validated measures of control strategy to current and future performance in complex environments--This intensive study of the relationship of control strategy to performance would permit the specification of optimal control strategies for important flying tasks. These could then be explicitly trained for.

- The need to devise causal models of validated measures of control strategy, to integrate these into HOPE, and to relate these to training procedures in complex environments--The dynamic model of control strategy and control dynamics (plant) learning which could be used to select the best training procedure for individuals in a given simulator, as well as providing an extremely valuable design aid for simulators. This research is the capstone for the effort begun in the present project.

REFERENCES

Alegria, J. Sequential effects of fore-period duration: some strategic factors in tasks involving time uncertainty. In P.M.A. Rabbit & S. Dornic (Eds.), Attention and performance V. New York: Academic Press, 1975.

Anderson, John R. Arguments concerning representations for mental imagery. Psychological Review, 1978, 85, 249-277.

Bertelson, P. The refractory period and choice reactions with regular and irregular interstimulus intervals. Acta Psychologica, 1967, 27, 35-56.

Engler, H. F., Davenport, E. L., Green, J., Sears, W. E. Human Operator Control Strategy Model, Final Report AFHRL-TR-79-60, Wright-Patterson AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory, April 1980. AD-A084 695.

Kahneman, D. Attention and effort. Englewood Cliffs, New Jersey: Prentice-Hall, 1973.

Kleinman, D. L., Baron, S. & Levison, W. H. An optimal control model of human response, Parts I & II. Automatica, 6, 357-83, 1970.

Knoop, P. A. Survey of human operator modeling techniques for measurement applications. AFHRL-TR-78-35, AD-A058 327. Wright-Patterson AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory, July 1978.

Obermayer, R., Swartz, W., & Muckler, F. Interaction of information displays with control system dynamics and course frequency in continuous tracking. Perceptual and Motor Skills, 1962, 15, 199-215.

Moray, N. A data base for theories of selective listening. In P.M.A. Rabbit & S. Dornic (Eds.), Attention and performance V. New York: Academic Press, 1975.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Posner, M. I. & Boies, S. J. Components of attention. Psychological Review, 1971, 78, 391-408.

Poulton, E. C. Tracking skill and manual control. New York: Academic Press, 1974.

Standards for Educational and Psychological Tests. New York: American Psychological Association, 1974.

Welford, A. T. Fundamentals of skill. London: Methuen & Company, Ltd., 1968.

Welford, A. T. The "psychological refractory period" and the timing of high-speed performance--a review and theory. British Journal of Psychology, 1952, 43, 2-19.

APPENDIX A

VALIDATION TEST EXIT INTERVIEW

1. Overall, how interesting did you find this task?

   1 = very dull
   2 = dull
   3 = OK
   4 = interesting
   5 = very interesting

2. a) What did you do to try to improve your performance?
   b) Can you explain in a little more detail?
   c) Anything else?

3. a) Thinking back over your total experience with this task, what do you think you learned about it that helped you improve your performance?
   b) Anything else?

4. a) While you were doing this task, did you pay any attention to the guidelines?
   b) Not at all? or, Can you tell me in a little more detail what you mean?

5. While you were doing this task, do you recall varying:

   a) how often you picked new positions for the position of the control stick? When did you vary this?
   b) the amount of deviation from the center you judged as acceptable? When did you vary this?
   c) how aggressively you reacted to excessive error? Did you vary this? When?

6. Would you say you varied your control strategy over the time you were doing this task? If necessary, (control strategy is a term we use to mean the way you do the task--what you look at, how you use the stick, how you judge your own performance, what you think about with respect to the task).

REMEMBER--For our study to be accurate, we need each person to come into the task "fresh." Please don't discuss your experiences with other students.

APPENDIX B

SUBJECT RESPONSES TO VALIDATION TEST EXIT INTERVIEW

In the material which follows, responses from all 28 subjects tested are grouped by similarity under each of the questions asked. For most of the questions, similar answers are grouped under an underlined label which is the authors' attempt to summarize the content of the responses. The underlined labels were not part of the interview.

Overall, how interesting did you find this task?

	Number Responses
Very dull	0
Dull	1
OK	9
Interesting	13
Very interesting	5

What did you do to try to improve your performance?

Changed Style of Control Stick Manipulation

used both hands for better control
find best way to hold bat handle
how was going to control stick
tried different hand positions
smooth out motions of stick and cursor
keep grip loose
pulled back for steadier movement
kept stick from moving unwillingly
took light touch
change hand positions (5 similar responses)
got used to stick
looked at shape of curve coming up, tried to simulate with wrist
learn response of joy stick
try not to over correct (4 similar responses)
tried not to move control sharply
tried to move smoothly
learned not to over move when there was little curve after there had been a lot of curve
tried not to react too quickly

Learned Control Dynamics

learned to relate stick position to cursor position
looking at center line
needed more movement when cursor was farther from screer
learned sensitivity between control stick and plus sign
got feel of stick, both physically and control dynamics
remember how quickly to shift control stick in the turns
relation between bat and cursor
tried to get acquainted with plus sign movement

Changed Level of Attention/Arousal

concentrate more (15 similar responses)
anticipate (3 similar responses)
relax (6 similar responses)
quit anticipating
brought eye drops for second day
gum chewing
singing

Memory

at points it wasn't a matter of thinking, you remembered how to handle the condition
remember previous tracks
same pattern repeated at the beginning was remembered
remember the road pattern
do better in place where had messed up before

Changed What was Attended To

keep eyes fixed on plus
paid more attention to going to left and had more trouble with that
fast enough that if you didn't pay attention it's easy to get messed up
would preview for a moment, then concentrate on few dots in front of plus, then repeat
concentrated on keeping it within guidelines
paid more attention with eyes to center line
concentrate, instead of sit back and relax
look ahead, see where curve is going to go
try to anticipate start to get on track as quickly as possible
predict curves
learned to pick out preview point to aim for
learn to look ahead
"drove" ahead of plus
look ahead
the quicker the plus was moving (side to side), the more preview used
when fast, used about 4 inches preview
started using preview, later just where cursor was
learned to be able to think about other things without hurting performance
watch track, not plus
tried different ways of looking at screen
tried point to point
focused on center line
yesterday just kept between lines, then today kept on center line

Other

fast curves were easier than straight parts
watch the spacing between the dots; large spacing means fast movement
if started well, did well; if started off track, never quite recovered
used rhythm of movement
learned to hold still when was only a little curve
keep plus on track
gauge swing needed in control to make cursor perform properly on curves on ½ Hz track
used guidelines, when got too close to them, move towards center

Thinking back over your total experience with this task, what do you think you learned about it that helped you improve your performance?

Learned Control Dynamics

controlling action
getting used to the system
not to jerk
learned how stick movement related to cursor movement
familiarity with control stick and how it related to cursor movements on screen
learned relation between bat and cursor
response of joystick
on sharp curves and long path, start slow, move quickly, slow down at end to prevent overshoot
had to let up before curve ended or would over-shoot corner
certain curves caused problems
on curves, hard to get used to acceleration and deceleration going in and out
speed needed to be moved for different curves
learned how fast to move stick with relation to degree of turn
learned how far to push stick for different types of 'S' curves
sensitivity of bat handle
learned moving stick fast got fast reaction
moving stick slow, then trying to speed up caused slower reactions
learned to push stick for sharp curves
had to move stick fast to keep up with sharp curves
on curves, move stick way over, then back to control
learned lag in stick response
stick sensitive
learned what type of movements required to match certain track slopes

Learned Control Stick Manipulation

kept hand loose
positioning the stick before the start of the trial
hold hand steady
learned not to over correct
got familiar with track
good grip
changed grip
find best position to hold

Learned Tracks

learned road pattern (6 similar responses)
learned there were two roads (5 similar responses)
having hard track in middle helped later runs, like a hard practice
different tracks
could recognize beginning of road (3 similar responses)
switching between roads, slow road easier (2 similar responses)
getting more familiar with apparatus
remember characteristics of track, what expected to do
certain areas were hard so tried to remember techniques in these areas

Learned Effective Level of Attention/Arousal

concentrate (2 similar responses)
be relaxed (4 similar responses)
when slow, keep plus right on the center line
if lean forward, focus attention easier
concentrate more when problem curves occur again

Learned Good Attention Strategy

learned to ignore preview available on whole screen
focus just on upcoming track
too much preview was distracting; use preview up to next curve
quit anticipating
the way the dots moved, showed task requirements, indicated speed of bat handle required
watch the spacing between the dots
focusing on a point to aim for
if off center, stay within guidelines
narrower guidelines gave better cue for center line

Other

incentive to try harder
planned ahead more
used angles of track
wished road was slower on curves
mind is able to react to situations more quickly after they have been seen before
more did it, got better
learned what to expect

While you were doing this task, did you pay any attention to the guidelines?

Yes, Used Guidelines

yes--used as a perspective for how far off the road I was
yes--more to guidelines than center
on sharp curves, guidelines were narrow, so staying between them kept on center
only helpful as measure of error
on fast track--paid less attention to center line, and paid more attention to keeping the plus equally between guidelines
yes--tried to follow center line, but guidelines let know how much messing up
when I tried to simulate shape of curve used guidelines
yes--got ideas about how far away from the center cursor was
thought that without guidelines it would be hard even to keep close to center
yes--sometimes
when got outside of them, knew there was large error
got idea of shape and characteristics of road
helped anticipation
yes--watched them a lot at first
tried more to stay within guidelines, especially on steep curves
getting upset when went outside of guidelines last few trials
tried to stay within guidelines
tried to stay within guidelines more than on center line
at least not go outside
guidelines got thinner on curves (visual distortion of width)
during fast sweeps of track, keep plus in center of guidelines
more attention to guidelines than center
easier to keep within guidelines
on straights between curves, more interested in keeping between guidelines than on center line
considered helpful
without guidelines, would've been more difficult
guidelines helped you see the curve
guidelines gave indication of how far off of the center you were
inside the guidelines is OK
yes--gave a better idea of where the thing was going
yes--if lost the center, would concentrate on staying inside the guidelines until could recover
easier to stay inside the guidelines
yes--when narrower, signalled need for quicker motion
guidelines helped judge error
thought guidelines were more important than center line, especially on sharp turns
yes--told not to cross guidelines made me want to stay inside them

No, Mainly Used Center

concentrated more on center line
mostly tried to stay on center
not much

every now and then, but mostly center of track
mainly concerned with center lines
didn't pay any attention to them
more interested in the center line
no--kept eye on center
didn't worry about guidelines
on slow--paid more attention to center
mostly watched the center, but on corners just tried to get within the guidelines
on center was more important than within guidelines
when slow, used center line
aware of their presence, but didn't use them to guide
paid more attention to center line
tried to stay on center line; sometimes between guidelines
if follow center line, tended to overreact more than if let the cursor stay on most convenient side
yes--main attention on center, guidelines provide boundary, use as such
first few times, tried to stay on center and overshot guidelines
not really
more or less kept attention on center of track, if guidelines not there would have done about the same

Other

resembles "electronic racing games"
expected roads to vary rate
too hard to keep on center line
as experience increased, got becter

While you were doing this task, do you recall varying how often you picked new positions for the position of the control stick? When did you vary this?

Yes, Reduced Variance Over Time

slowed down how often stick changed
learned that task didn't need much movement as it seemed through curves, more continuous
more gradually at end of last set
picked position and held it longer
learned it took less movement to correct
later moved less to stop jerkiness
after a while, don't have to move it quite as often

Yes, Increased Variance Over Time

varied quite a bit
started reacting faster, until hand got tired
varied faster later in learning
as sets went on, reactions were faster
started out picking new positions often later
moved more often
lot more fluid later on
varied more frequently later in learning

Yes, Moved More on Curves

at first, tendency to over-correct
after got used to it, moved continuously on curves
move stick more often on sharp curves
yes--move fast for steep slopes
left stick longer on straightaways, moved more on curves
yes--learned how to control stick to make proper movements on curves
curvier tracks--swift back and forth
on sharp curves moved stick faster
pushed stick over and held it for curves

Yes, Moved More on Straight Segments

on straightaways, had hard time controlling because stick is large
corrected often when the road was straight, not as often on curves
changed stick position more on straight parts

Yes, Moved Less on Straight Segments

for straighter sections used firm, very precise motions being careful not to move too much
on straighter segments left it still more
pausing movements on straights
didn't move much when little curve

Yes, Depended Upon Track

slower track--slow, careful movements
smooth movement when track moves back forth symmetrically
moved stick less when track was slow
start slow sometimes move slow and smooth, other times more quickly, depending on track characteristics
jerkier movements when the track varying little
with different tracks it did

No, Used Continuous Movement

mostly continuous motion
didn't change much except maybe at first
not too many times hold still
move a little bit, see what happened, then move again
mostly jerky movements
move a little bit at a time
felt making continuous slow motions, except on straights
mostly continuous movement (esp., on zig zag ½ Hz)
tried sweeping movements, but they tended to overshoot

Other

tried different bat handle groups
did notice some intermittent type movements
tried to find best grip to make fine adjustments
used guide lines to gauge how far to swing bat handle
more pausing motion in ½ Hz
wasn't aware of any changes
at end of last set, task seemed easier

While you were doing this task, do you recall varying the amount of deviation from the center you judged as acceptable? When did you vary this?

Yes

at first, within guidelines was good
definitely within guidelines
fast road--considered the guidelines acceptable
yes--didn't concentrate on being on dead center
later--found just staying within the guidelines increased score
yes--first couple of trials, inside guidelines acceptable
yes--at first tried to keep in guidelines
earlier--just tried to stay within guidelines
did concentrate on keeping between guidelines
considered an 1/8" acceptable on straights, but stay on center for.
beginning--midway between center and guidelines
in later trials, would make an effort to get back on center
got more stringent through task
later--would move back to get back on the track
as trial went on, more and more stringent requirements
at first, overshot corners, then got more careful
tried to decrease margin of error
stricter standard later in learning
slow road--tried to stay on center
when it was straight, tried to keep it closer
cut corners
allowed more error earlier
early--would wait for turns to catch up if got off track
starting--anything close was good
later--focused on center line
tried to keep on center, except on fast curves
yes--when diagonals got longer, sharper, closer together, not so worried about deviation
sometimes straightened out sharp curves
made compensations for times when thought it was too hard so that would do better in future
yes--corners got neater
if slightly off, would let it go in early trial
later stayed more on the center
later, tried for right on center
towards end--if there was noticeable space between cursor and center line, it was unacceptable
at end, center line was easier to follow
yes--as total error went down, tried to keep closer and closer
never deliberately let it stay off center unless felt that a small corrective control stick motion could not be made accurately

No

always tried to stay close to center
if went outside of guidelines, got "upset"
tried for the center line all the way through

then tried for the center line
trying for center
tried to stay as close to center as possible
at beginning tried to stay on center
tried to keep it on center
not much change
maybe
no
no--kept constant
didn't change throughout task
not really
consider ¼-inch as the maximum acceptable constant throughout
no--tried to keep it on the center

Other
curve
learned it was easier to stay close and hit a lot of points, as on a straightaway
if any part of cursor was tracking center, then error negligible
not possible to keep center of cursor on track
later--keep edge of cursor on road

While you were doing this task, do you recall varying how aggressively you reacted to excessive error? Did you vary this? When?

Yes

a lot at the beginning
initially overly aggressive, became more passive
at beginning overshot a lot, got better as it went along
at beginning--jerked more
more aggressive earlier
if outside lines, did to correct error fast
sometimes would overcompensate
early--when errors made, tried to correct fast
smooth it out
yes--got less jerky toward end
later learned to control it
reacted less aggressively later, partly due to frequent jump in cursor
lessened reaction to error to avoid overreaction
got more confident
at end--movements more gradual
became smoother later
later--tried to move more slowly to be better
more aggressive if cursor lagged more
yes--once got used to joystick, tried to correct errors more quickly
reacted faster later on in learning
got more aggressive
more aggressive later in learning
got more aggressive later on
later--would go back to road
toward the end of the sets reactions got quicker
towards end--got upset when got grossly off track or even near boundaries
more abrupt in movements to correct error
no--jerk it back to where it should be
try to correct error quickly
used different "perspective" techniques (toward end)

No

usually over-corrected through both days
not particularly
constant--reacted quickly
always reacted to stay between guidelines
no--reactions remained about same throughout
overshooting most all the time

Other

got frustrated
first--would wait for road
reacted quickly
when in error, often tried to correct that error rather than aim to have correct performance on upcoming track
concentrate more
went off track when movements sharp

overshoot on curves
hard to get back quickly without re-overshooting
quick sudden movements tended to cause bad errors
reacted too quickly
guideline overshoot on one side often lead to guideline overshoot on other side (overcorrected error)

Would you say you varied your control strategy over the time you were doing this task?

Yes

Screen
started looking more at the whole screen rather than just at cursor
pay more attention to screen than stick

Anticipate
anticipate
yes--first tried to anticipate the track
learned to look ahead of cursor instead of right where cursor was
yes--started looking ahead
generally, look ahead was to next curve
tried to anticipate start
tried to anticipate how much control for curves

Feel for the Stick
got feel for the stick
got feel for stick
relax on stick
change the way held the stick for better control
feel of the machine
smooth stick control

Yes--Trying to Improve
notice weak points
trying to improve
towards end it was like driving, more natural
better judgement of speed
tried to perfect to being on center
it developed
stricter judgement later on
concentrate more on lowering score
a little bit on each run to improve
yes--as got better

Reposition for Comfort
repositioned for comfort
get more comfortable position
relaxed later on

Concentration
tried to hold concentration
more concentration
concentrate on center on slow parts
yes--tried to keep attention on task
concentrated more on those later
later concentrated more on plus, then went back to preview tracking

Anticipation
went from using preview to not using it
looking too far ahead
went from anticipating to just following the spacing between the dots

Varied Strategy Between Tracks
learned to control differently in slow and fast parts
think of whole road rather than just center
yes--got used to curves narrow and right
noticed transfer
varied strategy between fast and slow portions of curve
be more fluid when fast
pay more attention on slow
tried to move more on the slow curves
started remembering past trial on same track

Other
at first--worried about center line
towards the end, knew if keeping within guidelines would score well
only thing that was tried was to figure out where it would start
yes
first keep in guidelines
fluid motions
slow down reactions
keep on center line
over-correcting to the left
at first got by between guidelines
stay close to line
not overcompensating
learned how to set angle of attack to keep it between the guidelines
yes--keeping in lines
easing into curves
by attacking problem, rather than letting problem take care of itself
when score went down little by little, thought scores might be fixed
thinking about experiment's purpose
had problems with straighter curves
yes--somewhat
yes--if it has to do with learning the control, varying the movements
to get certain cursor movements
little adjustments for straightaways

No Strategy
not appreciably
not sure if had a strategy
no
silly to call "strategy" for such an easy task

tried point to point
focused on center line
yesterday just kept between lines, then today kept on cente

dots

11

nes

self
ked

ents