3 1176 00159 6007

NASA CR-166,314

NASA CONTRACTOR REPORT 166314

NASA-CR-166314
19820014375

# Technical Approaches For Measurement of Human Errors

Warren F. Clement
Robert K. Heffley
Wayne F. Jewell
Duane T. McRuer

LIBRARY COPY

APR 22 1982

LANGLEY RESEARCH CENTER
LIBRARY, NASA
HAMPTON, VIRGINIA

CONTRACT NAS2-10400
May 1980

NASA

NF02307

NASA CONTRACTOR REPORT 166314

# Technical Approaches For Measurement of Human Errors

Warren F. Clement
Robert K. Heffley
Wayne F. Jewell
Duane T. McRuer

Systems Technology, Inc.
13766 South Hawthorne Blvd.
Hawthorne, California 90250

Prepared for
Ames Research Center
under Contract NAS2-10400

National Aeronautics and
Space Administration

**Ames Research Center**
Moffett Field, California 94035

N82-22249#

## ABSTRACT

Human error is a significant contributing factor in a very high proportion of civil transport, general avaition, and rotorcraft accidents. Finding ways to reduce the number and severity of human errors would thus appear to offer promise for a significant improvement in aviation safety. Human errors in aviation tend to be treated in terms of clinical and anecdotal descriptions, however, from which remedial measures are difficult to derive. Correction of the sources of human error requires that one attempt to reconstruct underlying and contributing causes of error from the circumstantial causes cited in official investigative reports. Relevant measurements based on a comprehensive analytical theory of the cause-effect relationships governing propagation of human error are indispensable to a reconstruction of the underlying and contributing causes. This report presents the technical details of a variety of proven approaches for the measurement of human errors in the context of the national airspace system. Primary emphasis is on unobtrusive measurements suitable for cockpit operations and procedures in part- or full-mission simulation. Procedure-, system performance-, and human operator-centered measurements are discussed as they apply to the manual control, communication, supervisory, and monitoring tasks which are relevant to aviation operations.

## FOREWORD

This report was prepared under NASA Contract NAS2-10400 sponsored by the Man-Vehicle Systems Research Division of Life Sciences at Ames Research Center. The contract technical monitor was Dr. David C. Nagel, the Systems Technology, Inc., (STI) technical director was Mr. Duane T. McRuer, and the STI project engineer was Mr. Warren F. Clement. This report was prepared during the interval from November 1979 through July 1980.

The authors have endeavored to acknowledge by textual reference the specific contributions of the many other researchers which have helped to make this work possible. In addition the authors wish to thank the following individuals for their assistance in preparing this volume: the contract technical monitor for his constructive criticism and guidance, and Sharon A. Duerksen of the STI staff for her help in publishing the volume itself.

TABLE OF CONTENTS


| Section | | Page |
|---|---|---|
| I | INTRODUCTION . . . . . | 1 |
| II | MEASUREMENTS FOR IDENTIFYING HUMAN ERROR . . . . . | 5 |
| | A. Types of Errors and Distinguishing Measurements . . | 5 |
| | B. Planning for Non-Intrusive Measurements in the Experimental Design . . . . . | 7 |
| | 1. Part-Mission Simulation . . . . . | 7 |
| | 2. Full-Mission Simulation . . . . . | 9 |
| | C. Steps in Planning for Measurements . . . . . | 10 |
| III | PROCEDURE-CENTERED EVALUATION . . . . . | 16 |
| IV | SYSTEM PERFORMANCE-CENTERED EVALUATION . . . . . | 25 |
| | A. Frequency Domain Measures . . . . . | 26 |
| | B. System Error Measures . . . . . | 28 |
| | 1. Individual Event Outcomes . . . . . | 28 |
| | 2. Ensembles of Event Outcomes . . . . . | 30 |
| | C. Measures of Safety and Operational Capability . . . | 31 |
| V | HUMAN OPERATOR-CENTERED EVALUATION . . . . . | 32 |
| | A. Psychomotor Behavior Identification Techniques . . . . . | 34 |
| | B. Approaches to Psychomotor Behavior Measurement . . . . . | 36 |
| | 1. A Taxonomy of Psychomotor Measurement . . . . . | 36 |
| | 2. Measurement Approaches Appropriate to Human Error . . . . . | 51 |
| | C. Measuring Human Response to a Change in the (Control) Task Situation . . . . . | 58 |
| | 1. Change in the Organization of Manual Control Actions . . . . . | 59 |
| | 2. Change in the Controlled Element . . . . . | 61 |
| | 3. Monitoring Manual Control Actions . . . . . | 62 |
| | 4. Monitoring Automatic Control . . . . . | 64 |
| | 5. Monitoring Manual and Automatic Control . . . . | 66 |

TABLE OF CONTENTS (Concluded)


| Section | | Page |
|---|---|---|
| | D. Evaluating Measures Which Reflect Operator Workload | 68 |
| | 1. Subjective Psychometric Ratings | 73 |
| | 2. Objective Workload Correlates | 73 |
| | 3. Psychophysiological Correlates | 81 |
| | E. Eye Point of Regard (EPR) Measurements | 82 |
| | 1. Background | 83 |
| | 2. Reduction of EPR Data | 86 |
| | 3. Future Applications | 88 |
| VI | CONCLUSIONS | 89 |
| REFERENCES | | 97 |

| Appendix | | Page |
|---|---|---|
| A | Section IV from STI TR-1156-1, Classification of the Sources and Distinguishing Characteristics of Error | A-1 |
| B | Adaptive Psychomotor and Cognitive Tasks for Measuring Excess Control Capacity | B-1 |

LIST OF TABLES


| Number | | Page |
|---|---|---|
| 1 | Measurements for Identifying Human Errors . . . . . | 6 |
| 2 | Steps in Planning for Measurements in the Experimental Design . . . . . | 12 |
| 3 | Mission Phase, Task, and Human Element Operations Breakdown for Approach and Landing . . . . . | 20 |
| 4 | Typical Performance Metrics . . . . . | 29 |
| 5 | Examples of Simulated Inputs Suitable for Psychomotor Measurements . . . . . | 47 |
| 6 | Types of Cognitive and Perceptual-Motor Workload Measurements . . . . . | 71 |
| 7 | The Original Cooper Scale . . . . . | 74 |
| 8 | Pilot Opinion Rating Scales . . . . . | 76 |
| 9 | Measurements for Identifying Human Error . . . . . | 90 |
| 10 | Quantitative and Qualitative Measurements and Evaluation Criteria . . . . . | 91 |

## LIST OF FIGURES


| Number | | Page |
|---|---|---|
| 1 | Time History on Final Approach with Flight Director HUD | 11 |
| 2 | A Procedural Block Diagram Culminating in System and Pilot-Centered Evaluation Measures | 18 |
| 3 | Sequence of Tasks Performed During Approach and Landing | 21 |
| 4 | Representative Example of a Scenario | 27 |
| 5 | The Identification Process | 37 |
| 6 | Structural Isomorphic Model of Man-Machine System | 40 |
| 7 | Translation of a Verbal Task Description to a Pilot-Vehicle Loop Structure (Straight Climb on Instruments) | 41 |
| 8 | Translation of a Verbal Task Description to a Pilot-Vehicle Loop Structure (Helicopter Approach to Hover) | 42 |
| 9 | Algorithmic (Linear Optimal Control) Model of Man-Machine System | 43 |
| 10 | Time History on Final Approach with Head-Down Flight Director | 56 |
| 11 | Time History on Final Approach Using Flight Path HUD with Flight Director | 57 |
| 12 | Simplified Sketch of Flight Path HUD with Flight Director (Longitudinal Axis Only) | 60 |
| 13 | Modified Cooper-Harper Rating Scale | 75 |
| 14 | Inverse Effective Time Delay as a Function of the Order of Lead Equalization Required of the Pilot | 80 |
| 15 | Effective Time Delay as a Function of Average Neuromuscular Tension | 80 |
| 16 | Measured Dwell Fractions and Transition Link Fractions for Manual and Automatic Approaches | 85 |

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AC | Advisory Circular issued by FAA |
| ADT | Auditory Discrimination Task employing a binary tone choice |
| ATC | Air traffic control |
| CNI | Communications, navigation, and identification |
| DFA | Describing Function Analyzer |
| EEG | Electroencephalograph |
| EMG | Electromyograph |
| EPR | Eye point-of-regard |
| FAA | Federal Aviation Administration |
| FFT | Finite Fourier transform |
| GSR | Galvanic skin resistance |
| HUD | Head-up display |
| ILS | Instrument Landing System |
| ITAE | Integral of the time-multiplied absolute error |
| LMM | Locator, middle marker |
| LOM | Locator, outer marker |
| NASA | National Aeronautics and Space Administration |
| NG | Not acceptable (Fig. 5) |
| NIPIP | Non-intrusive pilot identification program |
| OK | Acceptable (Fig. 5) |
| SID | Standard instrument departure |
| STAR | Standard terminal arrival route |

# SECTION I

# INTRODUCTION

Findings by the Flight Safety Foundation, the National Transportation Safety Board, and others indicate that human error is at least a major contributing factor in a very high proportion (80 percent or more) of civil transport, general aviation, and rotorcraft accidents. Finding ways to reduce the number and severity of human errors would thus appear to offer great promise for a significant reduction in accidents and improvements in aviation safety.

The proportional involvement of human errors in aviation accidents has been relatively stable in spite of many changes in the air traffic control system and typical cockpits. This does not necessarily mean that an irreducible minimum has been reached, however. Instead we appear to be on a plateau in understanding the quantitative details of just how the human elements contribute. To make a significant dent in error reduction requires a better appreciation for the sources and causes of human errors as they affect the total aeronautical transportation system structure.

Human errors in aviation tend to be treated in terms of clinical and anecdotal descriptions, however. For a more concrete identification of the sources of human error, one must strive to separate original underlying and contributing causes from the circumstantial causes cited in official investigative reports. Furthermore, if one is to attempt correction of the sources of human error, their cause-effect relationships must be better quantified and classified.

Meaningful quantification and classification requires a sound underlying and unifying foundation in terms of mathematical models which subsume existing evidence, permit the planning of experimental measurements, guide the interpretation of results, and serve as the basis for extrapolation of results to other circumstances. Reference 1 was prepared to fulfill this need for a sound foundation.

Reference 1 presents a validated analytical theory of input-output behavior of human operators involving manual control, communication, supervisory, and monitoring tasks which are relevant to aviation operations. This theory of behavior, both appropriate and inappropriate, provides an insightful basis for investigating, classifying, and quantifying the needed cause-effect relationships governing propagation of human error.

Based on the human error classification scheme, Ref. 1 identified sources and/or origins of human error in the context of human input-output behavior. The concepts were illustrated by a typical task analysis as an example of the approach required for identifying sources of human error among critical skills. In this report we now discuss the technical details of a variety of approaches for the measurement of human errors in the context of the national airspace system with primary emphasis on cockpit operations and procedures in part- or full-mission simulation. First, in Section II, the general types of measurements implied by the theory of human error are described. These, in general, are needed to identify and, in lesser or greater detail, to quantify the human's errors and error-free operations. Because realistic behavior depends so strongly on simulation system factors, the degree of simulation required is addressed next. Section II closes with suggestions for steps to follow in planning effective measurements to reveal human error in part- or full-mission simulation. Section III takes up in more detail specific aspects of the procedure-centered evaluation of human error based on the typical task analysis from Ref. 1. Section IV then elaborates on system-performance centered evaluation and Section V, on operator-centered evaluation of human error. Section VI concludes this report by summarizing the recommended measurements. References and two supporting appendices follow Section VI.

Let us begin with a commentary on the state of affairs regarding simulator measurements. In general, without focusing on human error, per se, the quantitative measurements which are routinely made during aircraft simulations are woefully inadequate or at best very limited in their scope. Seldom do measurements go beyond statistical manipulation

of certain basic vehicular states and controls or the mere gathering of time histories which reflect overall system performance. And usually the only measurements of any direct value to the experimenter are the pilot ratings or pilot commentary. What is nearly always lacking is a measurement which quantifies the actual or effective pilot behavior, i.e., the functional response to simulus, and a concise measurement which quantifies the overall man-machine system response latency or bandwidth in command following and disturbance regulation.

As a result of not having made effective simulator measurements, we still find ourselves not really knowing or understanding in clear quantitative terms how pilots fly aircaft, make decisions, cope with stress or workload, and develop skills within the context of the national airspace system. Without effective quantitative measures in these areas, it is therefore not possible to make quantitative measurements of human error — if one cannot quantify correct behavior, then one cannot quantify incorrect behavior.

It should be made clear that we are, in fact, capable of making effective measurements in the simulator environment. It is just not done comprehensively on a routine basis. Every simulation has its own very limited objective, and good measurements might be made in support of that objecive. But usually no measurements are made beyond that. This approach is not acceptable in viewing human error.

The philosophical view which is promoted in this report is to strive to use a wide variety of measurement techniques in connection with the NASA Ames Research Center Man-Vehicle Systems Research Simulator Facility. This is justified by the large time and resource investment in the full-mission simulation approach.

The trick is to make measurements sufficiently unobtrusive that they do not interfere with the experiment or the operation of the simulator facilities. This is probably the main reason for the popularity of routine statistical and pilot opinion measurements. In general, more sophisticated measurement techniques interfere with the subject in some way (e.g., many psychophysiological measures), impede progress of the

experiment (getting special paraphernalia ready for use), or require excessive computation or computational capacity (e.g., various parameter identification approaches). We shall be sensitive to these aspects in discussing various approaches.

Another underlying idea in much of what is presented concerns the timeliness of reduced data and measurements. Simply stated, it is better to evaluate results as they are generated than after the fact in order to

Detect and correct experimental flaws

Truncate or extend the period of simulation

Accelerate the reporting process

Debrief subjects more effectively

Establish the status of learning or training with more confidence

Discover unforeseen results earlier.

The approaches presented and discussed herein tend to support this basic notion.

## SECTION II

## MEASUREMENTS FOR IDENTIFYING HUMAN ERROR

Measurement of human errors requires identification procedures that will take into account such characteristics of human behavior as adaptation and learning. Through adaptation the human operator changes his behavior to achieve system performance in a new environment, whereas by learning he changes his behavior in successive encounters with the same environment. Because change itself accompanies human error, we therefore need measurements which help to identify sources and distinguishing characteristics of human error apart from adaptation and skill development. Such measurements must, in themselves, not alter the behavior which would otherwise be adopted, and are therefore additionally qualified as "non-intrusive" measurements.

Reference 1 recommends in Section IV a basis for the classification of the sources and distinguishing characteristics of human error. Section IV of Ref. 1 is reproduced herein as Appendix A for convenience in referring to the definitions, sources, and causes of human error which need to be identified. The classification scheme is founded on a theory of human error. This theory is designed to aid in planning, conducting, and interpreting research on the common sources of human error which may underlie the ostensible causes and factors given by the clinical lists in Section III of Ref. 2 and the anecdotal descriptions in Ref. 3.

### A. TYPES OF ERRORS AND DISTINGUISHING MEASUREMENTS

Prerequisite examination in Appendix A of the definitions, sources, and causes of human error which need to be identified leads us to suggest the preliminary arrangement of distinguishing measurements in Table 1 for further consideration herein. Notice that a particular measurement may be capable of identifying more than one type of error. For this reason interpretation of a variety of measurements of effects

TABLE 1

MEASUREMENTS FOR IDENTIFYING HUMAN ERROR

| Basic Sources of Error (from Ref. 1) | Groups of Distinguishing Measurements: System Performance-Centered Measurements: Location in state space and time with respect to authorized boundaries and schedules | Location in state space and time with respect to critical limits | Command-following or disturbance regulation bandwidth or latency | Stability | Operator-Centered Measurements: Describing functions | Remnant | Opinion ratings | Workload | Eye point of regard | Psychophysiological correlates | Location in state space and time with respect to acceptable limits | Procedure-Centered Measurements: Time line analysis of discrete stimuli, responses, sequences, and latencies in procedures |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Extreme command or disturbance amplitude | X | X | X | X | X | X | X | X | | X | X | |
| Extreme command or disturbance bandwidth | X | X | X | X | X | X | X | X | | X | X | |
| Controlled-element change | X | X | X | X | X | X | X | X | | X | X | |
| Reduced attentional field in single channel operations | | | X | | X | | X | | | | | X |
| Diverted or divided attention and perceptual scanning in multi-input operations | | | X | | X | X | X | X | X | X | | X |
| Reversals | | | | | | X | | | | | X | X |
| Illusions, kinetosis | | | X | | | X | X | | | X | X | |
| Spontaneous improper actions | | | | | X | X | | | | | X | X |

on the system of concern as well as on operator behavior may be required to identify a particular source or type of error in practice. In this respect the additional clues provided in Tables 9 through 11 in Appendix A may be especially useful in helping to interpret what we shall term system-centered and operator-centered measurements. Tables 12 and 13 in Appendix A are designed to assist in the even more difficult problem of identifying causes of error leading to inappropriate organization of perception and behavior at the executive level of the operator's activity-supervising control, transcending the (operator's) various directly involved systems such as the perceptual, cerebrospinal, autonomic and neuromuscular systems about the behavior of which particular measurements can be made.

## B. PLANNING FOR NON-INTRUSIVE MEASUREMENTS IN THE EXPERIMENTAL DESIGN

Experimental design must recognize beforehand the kinds of data interpretation which are desirable for identifying errors and should select the appropriate level of simulation, viz., either full-mission or part-mission in the case of the Man-Vehicle Systems Research Facility. Some advantages and disadvantages of each level of simulation, in part from Ref. 4, are offered here.

### 1. Part-Mission Simulation

Part-mission simulation offers economy by virtue of its ability to focus on a particular flight segment (e.g., letdown, approach, and landing) without spending simulator, crew, or experimenter time on portions of the flight (e.g., cruise) of lesser interest or in which fewer errors might be expected. Repeated simulation runs by one crew or an ensemble of simulations involving many crews become quite feasible.

The possibilities for improper execution of the myriad of normal and emergency procedures within a particular flight segment can be examined in more detail in advance, simply because the volume of alternative possibilities is reduced by comparison with that volume in full-mission

simulation. Thus one is more likely to be prepared with the necessary alternative details for more efficiently comparing and judging the discrete activities to detect procedural errors.

In a single run, procedures, behavior, and performance for all the tasks involved are characterized by specific concrete actions (or inactions) flowing in a sequence. Error is identified as an extreme deviation from a desired state. With many replications these concrete actions exhibit variability, either in kind or in degree. A probabilistic framework for particular events then becomes appropriate as a means of describing the experimental data. In addition, the potential tradeoffs (based on experience and training) involved in selecting various emergency actions can be exposed in the light of a utility concept (Ref. 1).

However, there are drawbacks in part-mission simulation. One of these drawbacks is associated with the influences of motivation, rehearsal, and skill development. Operator experience with each experimental situation must be controlled if meaningful comparisons are to be made. This, in turn, may compromise the realities of crew motivation. In addition, if each operator is to have experience with several types of controls and displays in sequence, the possibility of differences in performance depending upon which specific system was used immediately preceding must be considered. These carryover effects are particularly difficult to handle because no simple experimental or statistical technique exists for eliminating their influence on the results of a part-mission study.

Critics of part-mission simulation also like to cite the difficulty in establishing the validity of the pre-experimental environmental conditioning of the subjects, especially when terminal flight segments of a long term mission are involved. The identical elements theory of transfer of Thorndike will be cited to challenge the surrogate pre-experimental conditions which are required to induce fatigue, boredom, and complacency. This disadvantage can be more effectively countered by turning to full-mission simulation.

## 2. Full-Mission Simulation

The face validity of full-mission simulation, with its potential ability to duplicate the entire flight environment and the entire demand on the flight crew, is attractive and compelling because it offers an opportunity to capture the motivational subtleties residing in crew coordination and resource management which might contribute to human error and which might be overlooked (or not even duplicated) in part-mission simulation. Furthermore full-mission simulation would presumably allow the effects of fatigue, boredom, and complacency to exert a more realistic influence on vigilance and human error in terminal segments of the flight. These advantages were realized in the prototype full-mission simulation reported in Ref. 5.

Full-mission simulation is not without significant disadvantges, however. Reference 4 recognizes that crew training requirements are very substantial, especially if the cockpit procedures, controls, and displays being tested are not those to which the crew members are accustomed. For example, on-site flight instructors may be required to transition flight crews to an advanced technology cockpit prior to any full-mission simulation, if substantive errors are to be reduced to a level comparable with that toward which commercial air carriers are supposed to strive. Thus full-mission simulation of advanced technology operations implies a concomitant investment in air carrier crew transition training and certification, which can be very significant.

For procedure-centered human error data and other low probability events such as accidents, we can depend on full-mission simulation only for anecdotal and qualitative evaluation as in Ref. 5. Any statistical measures of confidence in procedural errors and other low probability outcomes would require months of accumulated experience at enormous cost. The outlook is much more favorable, however, for acquiring statistical measures of confidence in certain system-centered and operator-centered parameters from short-term temporal ensembles where the ergodic hypothesis is reasonably valid.

An example of the compact on-line efficiency with certain system-centered and operator-centered parameters can be measured from short-term temporal ensembles is shown in Fig. 1 for an approach and landing. In addition to the customary time histories of system state variables, system-centered measures such as bandwidth (exemplified by gain crossover frequency, $\omega_{c_{FD}}$), and relative stability (exemplified by phase margin, $\phi_{m_{FD}}$) provide time histories which can serve as event markers of changes occurring in the man-machine system. Furthermore, operator-centered measures such as the pilot's describing function [exemplified by amplitude, $\left|\hat{Y}_p\ (0.5j)\right|$, and phase angle, $\measuredangle\ \hat{Y}_p\ (0.5j)$, at a frequency of 0.5 rad/sec] provide time histories which can serve as event markers of behavioral changes.

## C. STEPS IN PLANNING FOR MEASUREMENTS

In order to provide a convenient checklist of some of the necessary prerequisites for careful planning of measurements, we have prepared the outline in Table 2. This table not only summarizes some of the discussion up to this point in the exposition but also serves as a reader's guide for the remaining sections of the report which emphasize procedure-centered evaluation in Section III, system performance-centered evaluation in Section IV, and human operator-centered evaluation in Section V.

Of particular importance is the deliberate emphasis in Table 2 on performing essential steps in the pre-experimental analysis. Planning data collection beforehand specifically for the anticipated data reduction and statistical analyses is a general requirement for studies of human behavior. A significant investment of time and effort beforehand will assure more productive results from the measurements obtained in the actual experiment. In addition to ensuring that the assumptions required for the analyses are met, consideration of the fiducial statistical tests provides guidance in deciding how much data to collect. In some cases, evaluation of the power of a proposed test for detecting expected differences may lead to abandoning a measurement or even abandoning the experiment!

① Time Window Full

② Middle Marker

③ Executing a Missed Approach

④ NIPIP Not Converged

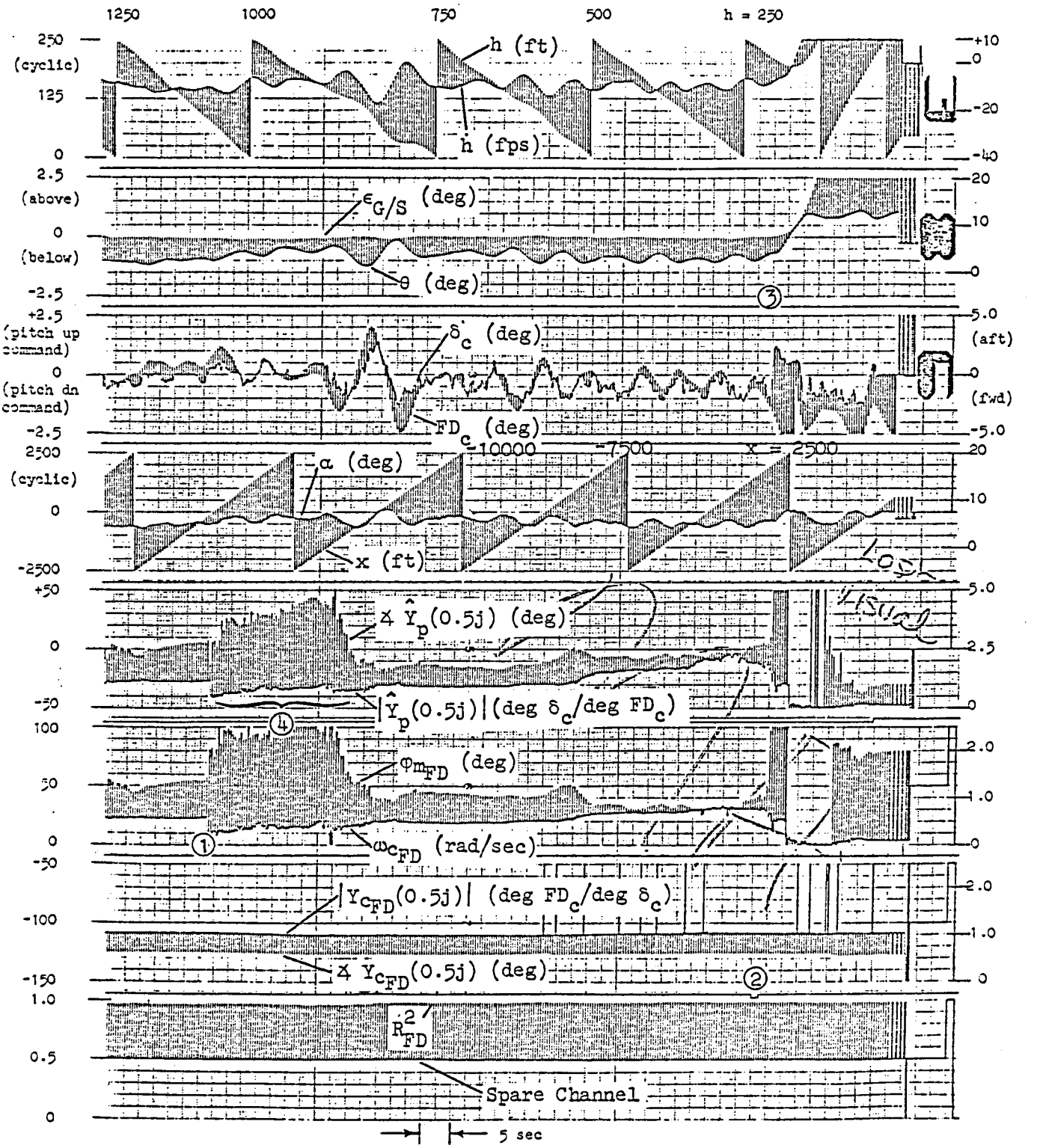



Figure 1. Time History on Final Approach with Flight Director HUD.

TABLE 2

STEPS IN PLANNING FOR MEASUREMENTS IN THE EXPERIMENTAL DESIGN

- Establish purpose, scope, and scenario
  - Elect part- or full-mission simulation
  - Specify mission phases, events, environment
  - Organize responsibilities, procedures, tasks for each crew member within each mission phase delineated by events
  - Specify inputs, types of activity (e.g., cognitive or psychomotor), outcomes and outputs associated with each task
- Perform essential pre-experimental analysis
  - Prepare activity time line analyses for normal and emergency operations together with likely alternatives for procedural errors which are foreseen
  - Classify non-intrusive measurements for the purpose of identifying errors
    - Procedure-centered evaluation based on time-sequences of all variables and events
    - System performance-centered evaluation
      - Command-following bandwidth or latency and critical exceedences
      - Disturbance regulation bandwidth or latency and critical exceedences
      - Safety; operational capability (distributions of state variables)
    - Human operator-centered evaluation
      - Pilot acceptance (distributions of state and control variables)
      - Temporal averages of task-specific dynamic behavior among crew members

TABLE 2 (Continued)

- Subjective ratings - appropriate workload indices for full-mission simulation
  - Objective workload correlates } Useful for part-task simulation
  - Psychophysiological correlates } Useful for part-task simulation

  (Note that objective workload correlates are useful for "calibrating" subjective ratings and psychophysiological correlates are useful event markers)
  - Eye point of regard: useful for event markers, temporal and ensemble distributions of attention
- Define measurement support and structure organization, and specify formats and media for output variables to be measured and recorded
  - Discrete outputs, events
  - Continuous signals to be sampled
  - Continuous signals without sampling
  - Closed-circuit video
  - Audio communications
  - Hard copy (e.g., subjective ratings and observers' notes)
- Estimate likely parameter values for proper and improper execution of activities within normal and emergency procedures
- Dry run portions of experiment and refine measurement techniques
- Specify output variables to be fitted by distributions from which probabilities can be estimated for the purpose of safety analysis verification and for interpretation in terms of decision analysis and workload analysis
- Manage and monitor data acquisition during experiment
  - Check against pre-experimental analysis
  - Look for measurement deficiencies
  - Keep up to date with as many on-line measurements as possible

TABLE 2 (Concluded)

- Relate measurements to commentary and observations

— Post experimental analysis

- Analyze interrelationships among
  - Procedure-centered measurements
  - System performance-centered measurements
  - Operator-centered measurements
- Identify or postulate sources of human error
- Perform planned statistical analyses (if any) and update hypotheses
- Refine behavioral models
- Recommend improvements to measurement procedures
- Organize and present results

— General recomendations

- Treat data as archival
- Acquire as much numerical definition as is practical (may be limited by storage and non-interference requirements)
- Do not restrict data acquisition to the narrow objectives of the experiment; it may serve someone else 10 years hence!

A final point in the design of experiments for studying human error in using controls and displays involves a logical problem that is restrictive and, perhaps for that reason, frequently overlooked. When control-display systems being compared differ in several characteristics, there is no possible way to conduct a single experiment and draw valid conclusions about which of the several differences in the controls and displays is responsible for any observed differences in system or operator performance. All that may be concluded is that the collection of differences in control-display design resulted in differences in performance. Identification of a single feature of a design as responsible for a difference requires measurements with systems in which only the single feature of interest is changed.

## SECTION III

## PROCEDURE-CENTERED EVALUATION

In Ref. 2 we noted the numerical predominance in Table 1 of procedural, supervisorial, planning, and communication/navigation/ identification (CNI) errors, which also appeared among the last four entries in Table 4. In-cockpit procedures; supervision of checklists; ATC clearance, communication, and bookkeeping; navaid selection; use of changeover points; and reporting to ATC for navigation on various airways/ route systems occupy a significant portion of the pilots' time, especially in areas of higher traffic density. Most of the errors identified by Ref. 5 are in these categories. Reference 5 has already cited the problems of handling the inordinate volume of documentation required in the cockpit to support these types of activities. Just handling this library in the cockpit is a monumental task, notwithstanding the responsibility for complete familiarity with an incredible array of procedures. These problems are compounded by the inefficiency of voice communication among crew members within the flight deck as well as between the flight deck and the ground facilities having jurisdiction over the flight. This inefficiency may lead to procedural errors and temporal latencies in discrete events and in stimulus-response relationships involving not only cognitive processes but also more than one human operator. Consequently we have adopted the suggestion that "'slips' at the precognitive level, either from faulty activation of schemata or faulty triggering of active schemata," may also be an implicit source of error underlying many of the cited causes which involve a procedural error as well as a flying error, even though "spontaneous improper action" appears explicitly in Table 1 of Ref. 2 only with rank 10(a) and in Table 4, not at all.

Measurement techniques are well-developed for identifying spontaneous improper actions, provided the sequences of tasks and actions necessary for mission success and failure have been thoroughly planned

and defined at the outset of an experiment. Such careful pre-experimental identification of procedures, both proper and improper, provides a framework exemplifying the spatial-temporal facets of the mission phase event- or time-line which are essential to the recognition and interpretation of "slips" at the precognitive level of operational behavior. The necessarily thorough pre-experimental definition of procedures was applied in Refs. 4 and 5, but in Ref. 5 the details of recording discrete actions such as setting switches or levers, responding to check lists, or coping with emergencies were relegated to an observer's commenting on a voice recorder, coupled with voice records of all flight deck communications. Since retrieval of "slips" from voice records is both tedious and cumbersome, as well as subject to the additional interpretation of the observer and participant, it is preferable to institute automatic recording of discrete actions by the crew members wherever possible. Thereafter to detect "slips" it is possible to employ automatic comparison of the recorded time-line of discrete actions with the pre-experimentally recorded time-line of "normal" and "emergency" procedures established for the scenario.

Our starting point for establishing a time-line of "normal" and "emergency" procedures for the scenario is the vehicle operational profile (or mission profile). To accomplish this essential pre-experimental planning, the mission is first defined and partitioned into a hierarchy of constituents. The primary constituents are mission phases. These are of a size and duration which allow the broadest factors (e.g., environmental variables) that influence human behavior to be identified. For example, if the mission phase be "approach and landing," our starting point is represented by Block 1 in the procedural diagram, Fig. 2. From this point of departure three categories of variables must be determined, viz., the procedural variables (i.e., the functions to be performed) in Block (7), the task variables in Block (8), and the environmental variables in Block (9), all of which exert an impact on the inputs to the man-machine system of concern. (We shall defer consideration of Block (2) to Section IV, where we discuss system-centered evaluation.) At the next level are the tasks, per se, in

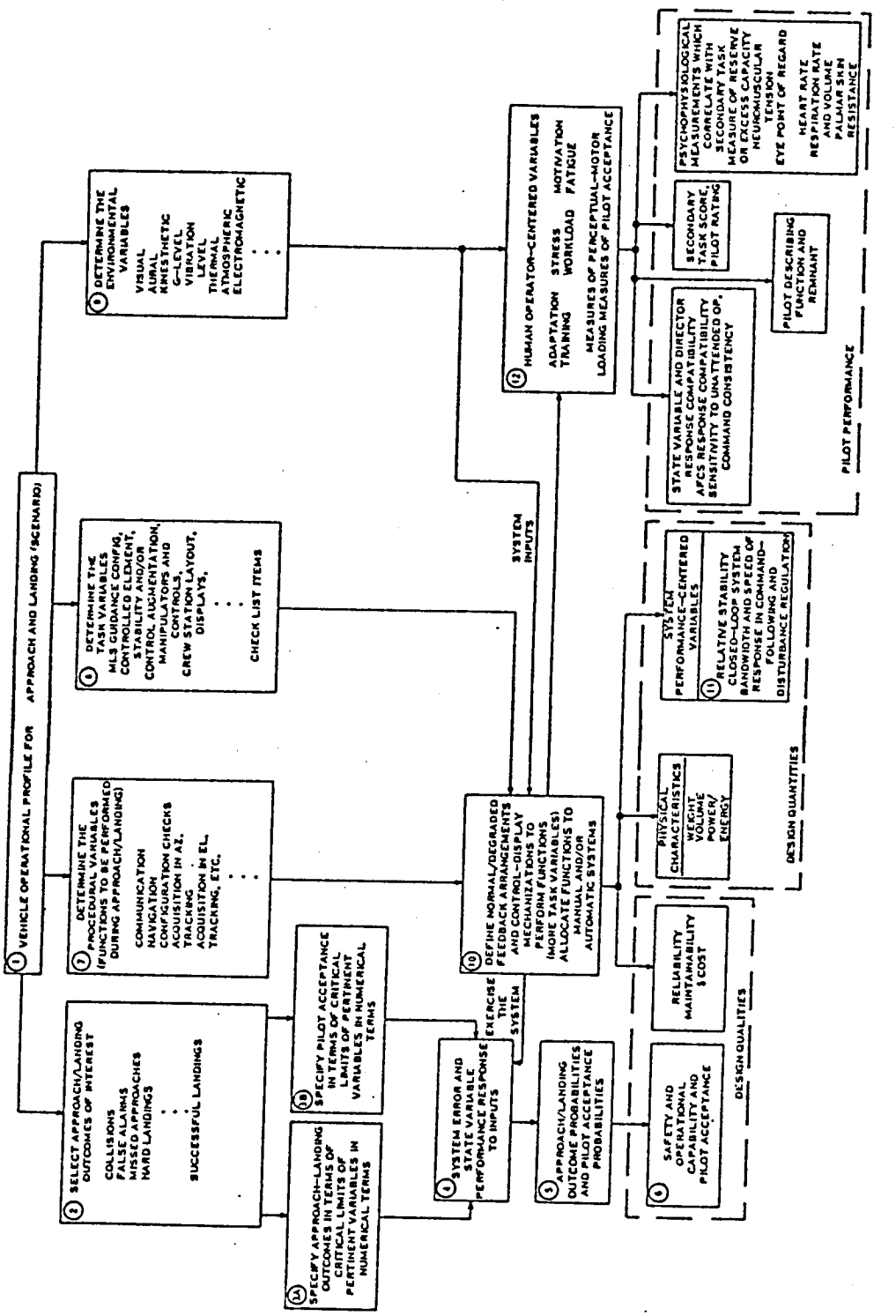



Figure 2. A Procedural Block Diagram Culminating in System and Pilot-Centered Evaluation Measures.

Block (10), which are associated with a particular operation in a sequence and are sized to permit the identification of "critical" skills. Aberrations in the execution of these skills ultimately determine the sources of contributions to human error.

A mission phase may be broken down into various subdivisions depending on its complexity. For our purposes here, we are ultimately interested in the elemental unit of all phases involving the human operator, the task. As a working definition here, we will define a task as an activity at the functional interface of the human operator and the objects and environments with which he interacts (Adapted from Ref. 6). We will further specify a task, for our purposes here, as a goal or criterion-oriented work increment involving application of a skill or set of skills by the human operator. Thus, by partitioning the mission phases into tasks, we can then identify those fundamental human operator behavioral factors, skills, which influence flight safety. For tasks which are critical to flight safety (i.e., exert a predominant influence in some sense), it is the proficiency with which a skill or set of skills is applied that we wish to consider in order to identify the underlying sources of human error.

To illustrate these remarks, Table 3 and its companion Fig. 3 (from Ref. 1) present an exemplary task breakdown for the pre-approach, approach, and landing mission phases of a Category 1 or 2 instrument approach. The tasks include checklists, tuning radios, requesting and receiving clearances, navigating as required by ATC procedures, etc., as well as flying the airplane. Each task is listed as an item in an ordered, nominal sequence. Conceivably this order might be changed or omitted in off-nominal circumstances, and this by itself may be a cause of error. Otherwise, no consequence of an erroneous execution of a task is explicitly indicated on the list.

Associated with each task are input and output modalities for the pilot (or other active crewmember). And, finally, with each task is an indication of the human behavior characteristics nominally involved in carrying out the task at hand. In many cases the nominal behavioral

# TABLE 3

## MISSION PHASE, TASK, AND HUMAN ELEMENT OPERATIONS BREAKDOWN FOR APPROACH AND LANDING

| PHASE OF FLIGHT | LOCATION ON FIGURE | TASKS | MODALITIES | NORMAL OPERATIONS |
|---|---|---|---|---|
| Preliminary preparations for approach | A | Request/receive approach clearance<br>Complete preliminary before-landing checklist<br>Check that all systems are operating (no flags)<br>Tune and identify navigation receivers to ILS<br>Tune and identify ADF's to LOM (LMM)<br>Preselect and enter subsequent communications frequencies<br>Set marker beacon switches and test<br>Set decision height on radio altimeter<br>Set inbound ILS localizer heading on respective course indicators<br>Maneuver to proceed to final approach fix | Verbal/Verbal<br>Visual/Manual/Store<br>Visual - Store<br>Visual/Manual<br>Visual/Manual<br>Visual/Manual<br>Visual/Manual<br>Visual/Manual<br>Visual/Manual<br>Visual/Manual | Precognitive<br>Precognitive; compensatory<br>Pursuit<br>Pursuit<br>Precognitive (if switchboard)<br>Precognitive; compensatory<br>Pursuit<br>Pursuit<br>Precognitive; compensatory |
| Initiation of lateral guidance acquisition | B | Maintain altitude<br>If procedure turn required select appropriate heading<br>Accomplish procedure turn<br>Report procedure turn inbound | Visual/Manual<br>Visual-Store<br>Visual/Manual<br>Verbal | Compensatory<br>Precognitive; compensatory |
| Preparations for acquiring vertical guidance | C | Acquire initial approach airspeed<br>Set partial flaps | Visual/Manual<br>Visual/Manual | Pursuit<br>Precognitive; pursuit |
| | D-E | Descend to (and maintain) initial approach altitude | Visual/Manual | Pursuit |
| | E | Set speed command system to desired speed<br>Increase flaps and reduce speed<br>Check missed approach procedure, decision height, and RVR | Visual/Manual<br>Visual/Manual<br>Visual - Store | Precognitive; pursuit<br>Precognitive; compensatory |
| Acquisition of lateral guidance | F | Initiate capture of localizer beam | Visual/Manual | Precognitive |
| | G-J | Stabilize on lateral flight path | Visual/Manual | Compensatory |
| | G-Q | Maintain lateral guidance | Visual/Manual | Compensatory |
| Acquisition of vertical guidance and completion of preparations for landing | H | Lower landing gear | Visual/Manual | Precognitive; compensatory |
| | I | Lower nose flaps and start bleeding more airspeed | Visual/Manual | Precognitive; compensatory |
| | J | Check time at outer marker<br>Capture glide slope beam — extend full flaps, acquire final approach airspeed, and establish sink rate<br>With safe landing gear indication, complete "final checklist"<br>Change to tower frequency<br>Report OM inbound | Visual - Store<br>Visual/Manual<br>Visual/Manual/Store<br>Visual/Manual<br>Verbal | Pursuit<br>Precognitive; compensatory<br>Precognitive; pursuit |
| | J-K | Stabilize on vertical flight path | Visual/Manual | Compensatory (INC); Pursuit (VNC) |
| | J-M | Maintain vertical guidance | Visual/Manual | Compensatory (INC); Pursuit (VNC) |
| Final approach | K-M | Maintain stabilized flight path (in all axes) | Visual/Manual | Compensatory (INC); Pursuit (VNC) |
| | L-M | Use extended glide slope or Category II beam for vertical guidance | Visual/Manual | Compensatory |
| Decision Height | M | Execute missed approach if required | Visual/Manual | Precognitive; compensatory |
| Flare | M-P | Reduce sink rate | Visual/Manual | Precognitive; pursuit |
| | N-P | Decrab to align airplane with runway | Visual/Manual | Pursuit |
| Touchdown and rollout | P | Contact with ground | Motion, Visual - Store | |
| | P-Q | Steer throughout rollout<br>Decelerate to a stop | Visual/Manual<br>Visual/Manual | Pursuit<br>Pursuit |

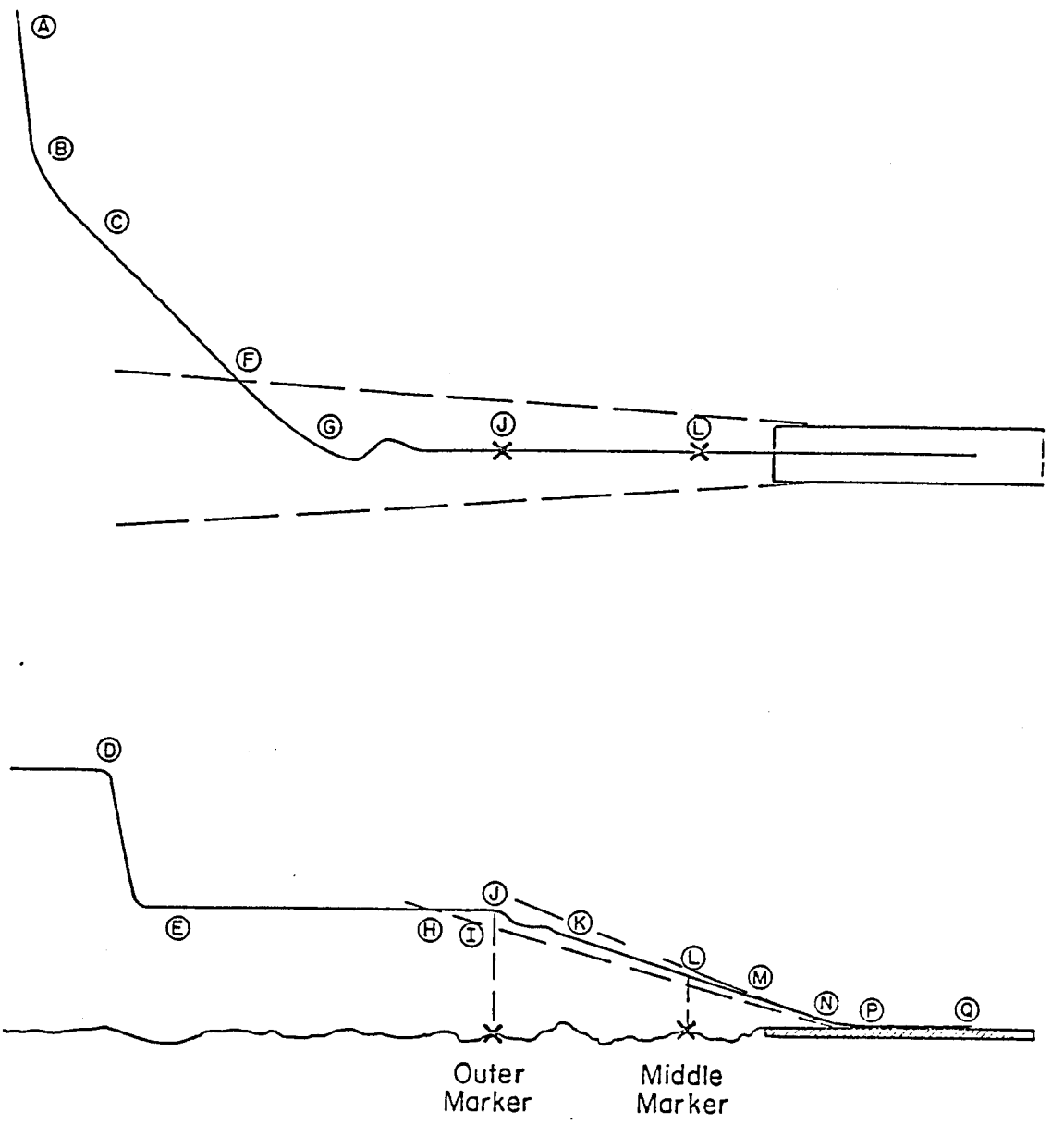



Figure 3. Sequence of Tasks Performed During Approach and Landing

characteristics may not be exhibited by actual crews, and this abnormal behavior may result in an out-of-tolerance system error.

In most of the tasks where precognitive operations are cited in Table 3 as nominal or customary, additional qualification is necessary. Such open-loop operations are normally of limited duration and are properly interspersed or concluded with closed-loop operations either directly, as in dual mode continuous control, or indirectly in the context of the off-line supervisory monitor described in Fig. 10a of Ref. 1. Omission of the closed-loop monitoring activity may in fact lead to human error as shown in Ref. 7. Examples are: tuning communications, navigation, and identification (CNI) equipment; selecting partial flaps; lowering gear; setting throttles; dumping fuel; and accepting ATC clearances which are either physically impossible or unsafe. To emphasize this point, some of the precognitive operations in Table 3 are accompanied by compensatory operations. The nature of the control and display interface with CNI equipment in particular will also determine whether channel frequency selection can be purely precognitive or must include compensatory verification.

For the measurement of human error, the nominal task breakdown illustrated here must be further subdivided to account for all possible outcomes. This is illustrated in Section V of Ref. 1 for the terminal end of the approach and landing mission phases. Other off-nominal aspects which should be considered are the accumulation of stress and degradation of skill. Each mission phase presents a combination of environmental and task stresses on the crew, and these stresses influence crew performance. After lapses in operational practice or in long duration flights, crew members have to cope with the problem of maintaining proficiency of skills which may be critical to flight safety. Skills performed infrequently prior to or during each flight, for whatever reason, are most likely to fall into this category. Of these skills, those having high workload factors by virtue of being time-constrained or because they involve complex operations are most likely to cause serious performance decrements. Several conditions may contribute to the degradation of these skills:

1) Lack of practice.

2) Inability to practice in the appropriate environment.

3) Interference or negative transfer arising from the practice of competing skills.

4) Physiological deconditioning due to fatigue induced by the environment or due to alcohol or drug stresses.

The tasks which are most likely to be affected by these human conditions should be especially flagged.

Most of the points made above have an intuitive appeal as well as a logical structure. This overall structure has been outlined here to provide an example showing the tying-together of elements into a whole which provides the necessary pre-experimental identification of normal and emergency procedures, both proper and improper. These procedures, in turn, provide a basis for identifying human errors among the recorded time-lines of discrete actions by crew members in a full mission simulation. Nevertheless, a word of caution is in order about the use and abuse of pre-experimental time-line analyses, which can be carried to the point of diminishing returns. For example, it is customary to estimate latencies and operator task "loading" from procedural time-line analyses. Conventional time-line analyses for estimating latencies and workloads suffer from several shortcomings. Accurate estimates of times required for the intangible elements of activities such as direction of attention, memory, and decision making are generally not available, and even the vague estimates are generally based on textbook descriptions of operator behavior in performing discrete tasks. But flight safety is not necessarily a function of operator performance as described in textbooks. Catastrophic events are precipitates of the interaction of very rare events (external and/or psychological) that may coincide capriciously in time. One cannot necessarily list the tasks required sequentially of an operator, add time allotments up to the 99th percentile, and show thereby that the job can be done acceptably. Instead,

discrete outputs and events usually provide the most useful benchmarks for establishing on-line measures of decision-making behavior. To establish a system latency (time for an "input" to propagate through the multioperator system) the "input" may, in the case of cognitive tasks, have to be considered to be present if and only if all of the information (based on continuous and discrete signals available) which is needed to derive the "input" as a conclusion to be acted on is present.

When the realities of pilot behavior under boredom or high stress are included, plus the contingencies in task requirements that depend upon prior timely execution of related tasks, the cost and complexity of extremely detailed pre-experimental task analyses may become unreasonable. Notwithstanding this word of caution, at least the level of detail illustrated in Table 3 will be necessary in order to detect procedural errors by comparison with a recorded sequences of discrete actions among crew members.

## SECTION IV

## SYSTEM PERFORMANCE-CENTERED EVALUATION

System performance-centered measurements can be divided into two categories: those which reflect design quantities and those which reflect design qualities. Design quantities include the dynamic system performance (relative stability, accuracy, closed-loop bandwidth or speed of response in command-following, and disturbance regulation) as well as the physical characteristics of the system. Design qualities may also be quantified and include safety, pilot acceptance, operational capability or effectiveness, reliability, maintainability, and cost. These measurements apply to automatically controlled aircraft and their subsystems; to control-display subsystems involving one or more human operators; and to communications, command, and control systems involving two or more operators. However, because a single measure cannot quantify both system and operator performance and because we are unable to express either operator acceptance or the reliability of the human operator in terms commensurate with the design qualities of equipment, it is necessary to introduce a variety of related qualities that characterize human operator compatibility, e.g., behavior adaptation*, learning, workload, stress, fatigue, motivation, and pilot opinion rating. An "optimum" system is one that has some "best" combination of all of these features.

---

* Through adaptation the human operator changes his behavior to achieve system performance in a new environment, whereas by learning he changes his behavior in successive encounters with the same environment. In terms of pilot behavior the improvement of system performance implies reduced effective time delay; reduced pilot-induced noise insertion (unwanted control action); increased allowable range of pilot gain variation consistent with closed-loop system stability; progression above the compensatory level in the successive organization of perception through skill development; and reduced workload to a level where the pilot is efficiently and gainfully occupied, yet able to cope to a prescribed degree with the unexpected. We shall devote Section V to a consideration of measurements which reflect human operator-centered evaluation in more detail.

Our starting point for establishing system performance-centered measurements is a vehicle operational profile. For the example of approach and landing, this is represented by Block (1) at the top of Fig. 2. Examples of operational profiles are given in Fig. 1, Table 3, Fig. 3 (from Ref. 1), and Fig. 4 (from Ref. 114).

## A. FREQUENCY DOMAIN MEASURES

Based on the operational profile, we have already noted in Section III the need to determine the task variables (Block (8) ) and procedural variables (Block (7) ) which support communications, navigation, identification, command (guidance), and control functions (Block (10) ) required by each phase of the scenario for normal and degraded operations. The design requirements for these functions are in turn dictated by four needs the first three of which are conveniently characterized by frequency domain measurements:

- Stability
- Command-following bandwidth
- Disturbance regulation bandwidth

} Block (11)

- Compatibility with the human operator — Block (12)

The satisfaction of these needs leads to the selection, sensing, shaping, and relative weighting of appropriate feedbacks in Step (10) and to their partition between manual and automatic systems. The relative degree of stability can be characterized by measuring phase or gain margins of stability, the closed-loop system bandwidth, and speed (or latency) of response in command-following and disturbance regulation. The relative ease with which phase margin and system bandwidth measurements can be made is illustrated by their respective time histories identified during the simulated approach recorded in Fig. 1. These measures are fundamental to any closed-loop system and are independent of whether control is automatic or partitioned among several human

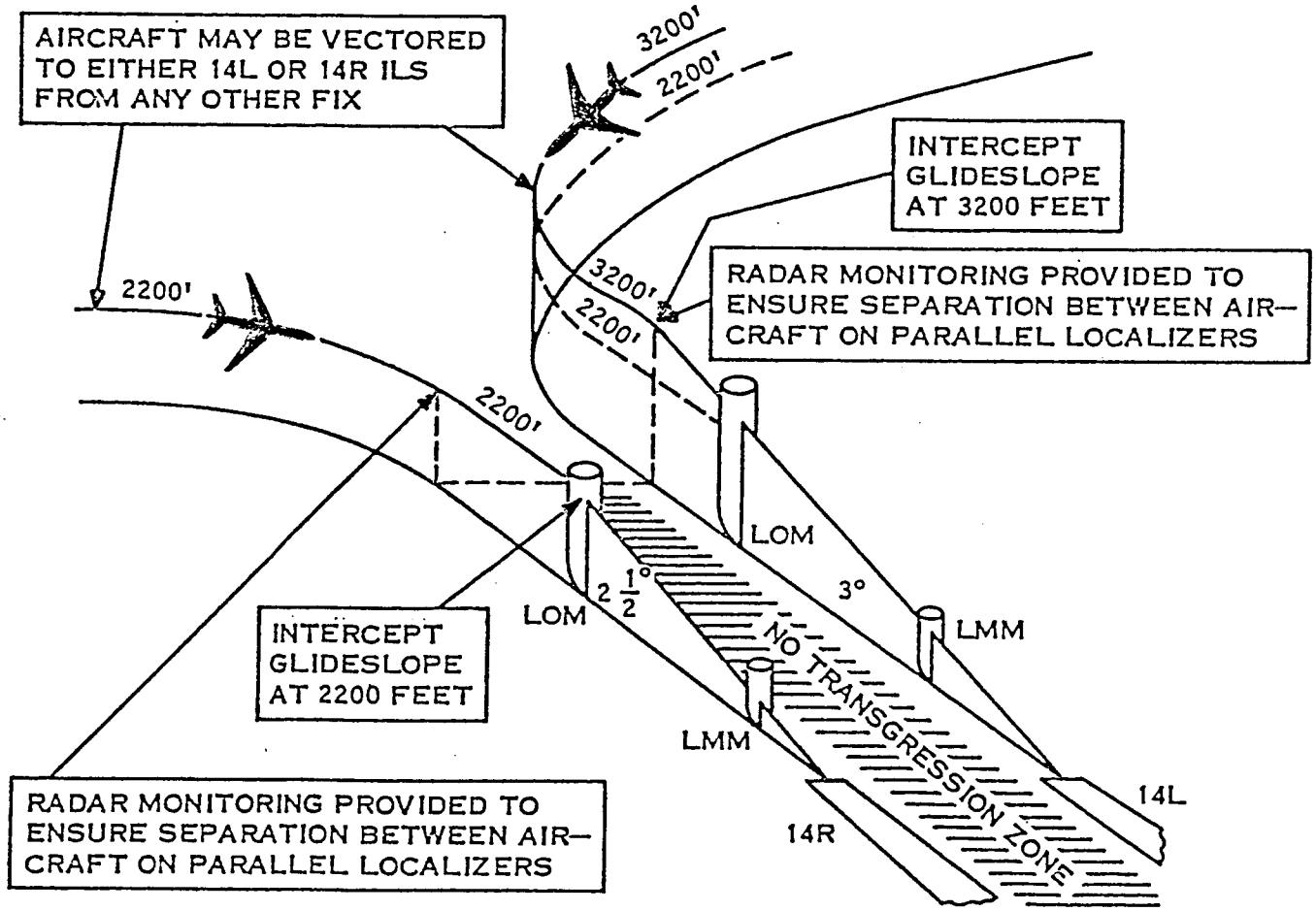



Figure 4. Representative Example of a Scenario (From Ref 114)

operators. For example, when measuring system bandwidth or system latency*, and cognitive tasks are involved, the "input" may have to be considered to be present only when all of its necessary constituent information is present, because the "input" may then and only then derive as a conclusion to be acted on.

## B. SYSTEM ERROR MEASURES

### 1. Individual Event Outcomes

Based on the operational profile, a list of the outcomes of the approach and landing phases of flight is also developed. This step is represented by Block (2) in Fig. 2, and a sample is provided by Table 15 in Ref. 1 and by Table 4 herein. Typical values for the critical limits for a subsonic jet transport are given in the appendix to Ref. 8. Analogous limits for a STOL aircraft are given on p. 115 in Ref. 9. The critical limits, in turn, are based on data from a variety of sources such as FAA Advisory Circular 20-57 on Automatic Landing Systems, applicable flight handbooks, and aircraft geometric, structural, and aerodynamic limits. Other limits that reflect acceptability of the approach and landing can be incorporated in similar evaluation criteria. For example, Ref. 10 suggests criteria for judging measured attitude and heading angles and normal acceleration on automatically controlled approaches. Margin from stall and the maximum rate of descent, both of which become more critical to the pilot as the approach angle steepens, should also be measured. Ref. 11 suggests analogous criteria for judging measured control displacements and rates. Critical limits such as these are represented in Bock (3B) in Fig. 2. System performance is examined in Block (4) . Proceeding in this manner, we can express in commensurate terms the performance criteria by which accomplishment of

---

* Latency = time for an "input" to propagate through the (multioperator) system.

TABLE 4

TYPICAL PERFORMANCE METRICS

| MISSION SEGMENT | PRINCIPAL FORCING FUNCTIONS | PERFORMANCE METRIC |
|---|---|---|
| Transition to missed approach configuration or to engine-out takeoff/ climb configuration | Configuration changes to establish trimmed transition flight path, path command, gusts | Settling time, ITAE, rms motion variables, probability of exceeding control limits, pilot activity (control axis crossings) |
| Timed approaches from a holding fix to parallel runways (see Fig. 4) | Leader's maneuvers, gusts, terrain, potential lateral and vertical conflicts | RMS deviation from desired position, probability of collision, exceeding control limits, or striking the ground, pilot activity (control axis crossings) |
| ILS approach | Beam bends and glide slope scalloping, gusts, wind shear | Settling time, ITAE (for beam capture), rms motion variables, pilot activity (control axis crossings), probability of exceeding limits on position and sink rate at terminal time |

A definitive treatment system performance criteria in both the time and frequency domains is given in Refs. 122 and 123.

the approach and landing (or other mission segment) will be measured and the penalties associated with errors in system performance.

Having specified critical limits for the approach/landing outcomes, we can determine the individual approach/landing outcomes by comparing the measured values of the pertinent state variables with their corresponding critical limits which represent system performance criteria. The steps required are represented by Blocks (3A) and (4) in Fig. 2.

As an example, in Ref. 8 a wind-shear model is used to determine the quantitative relationship between acceptable mean deviations (glide slope, d, localizer, y, and airspeed, $u_a$) at 100-ft altitude and at touchdown. These relationships are windows in "state space" that have the dimensions ($d_{100}$, $\dot{d}_{100}$, $u_{a_{100}}$) and ($y_{100}$, $\dot{y}_{100}$), respectively, for the longitudinal and lateral situations.

## 2. Ensembles of Event Outcomes

Proceeding in this manner we can alternatively express in commensurate probabilistic terms the performance criteria by which accomplishment of the approach and landing (or other mission segment) will be measured and the penalties associated with errors in system performance. Having specified critical limits for the event outcomes in Block (3A) (and (3B) ), we can compute in Block (4) amplitude and frequency distributions of ensembles of state variables and control variables which define the outcomes of interest. From these distributions outcome probabilities can be inferred. The results represented in Block (5) are probabilistic measures of the approach and landing outcomes such as those in Table 3 and of system acceptance in terms of attitude and heading deviations from trimmed values, normal accelerations, and control displacements and rates.

## C. MEASURES OF SAFETY AND OPERATIONAL CAPABILITY

Finally in the step represented by Block (6) in Fig. 2, the results from Block (5) are used to compute measures of safety and operational capability such as the expected number of approaches required to land, given an arrival in the terminal area (safety); the expected number of accidents, given an arrival (risk); and the minimum average time between landings (operational capability or effectiveness).

## SECTION V

## HUMAN OPERATOR-CENTERED EVALUATION

As in the case of system performance evaluation, our recommended starting point for establishing operator-centered evaluation criteria and corresponding measurements is a vehicle operational profile. For purpose of illustration, consider again the example for approach and landing in Table 3 and Fig. 3. The approach and landing profile includes several distinctly different classes of procedural variables as represented in Block (7) in Fig. 2. These include (i) visual-motor tracking of guidance references (flight control); (ii) discrete tasks, such as following checklists, making configuration changes (e.g., flap and gear extension), and routine communication, navigation, and identification (CNI) tasks; (iii) decision-making CNI tasks such as responding to ATC advisories or intrusions, and failure management such as coping with a partial loss of propulsion, compensating for a failed yaw-damper, or deciding to take over manual control of an axis; or (iv) the use of other perceptual-motor modalities such as verbally calling out altitudes during an approach. The diverse examples cited illustrate that there is no single type of operator-centered evaluation criterion and measurement which covers all of the operating procedures.

Measures of system performance, safety, and operational capability, coupled with other design qualities such as cost, reliability, and maintainability, might be sufficient for evaluating a completely automatic system. However for a piloted system, experience has shown that many other factors are involved in the ultimate assessment of sources of error. This is because measures of <u>system</u> performance, safety, or operational capability are insufficient for measuring <u>pilot</u> performance. For example, among different approach course tracking systems, the pilot may adapt his behavior so that an overall system performance measure remains relatively invariant and, therefore, unsuitable for inferring anything about pilot performance. Consequently, it is necessary to recognize and attempt to measure the operator-centered variables that

reflect sources of human error, namely, adaptability, learning, perceptual-motor workload, stress, fatigue, motivation, and pilot opinion rating (Block (12) in Fig. 2). Because these operator-centered variables are so important, we shall discuss psychomotor behavior techniques in the following subtopic (A) and present a taxonomy of psychomotor behavior measurement in the next subtopic (B) and follow that with a discussion of measuring human response to a change in the task in Subtopic C.

Nevertheless, some system performance measures are important factors in pilot acceptance and may thereby contribute to errors in judgment. These include variances of attitude, attitude rate, load factor, and control activity. Any of these, if too large, will lead to some degree of pilot dissatisfaction and possibly even pilot error.

Another system performance consideration related to pilot acceptance and errors in judgment is the harmony between manual and automatic control for systems that can operate in both modes, but in somewhat different manners. For example, in aircraft equipped with direct lift control or a collective control, the automatic system may conduct the landing maneuvers in a different fashion from the pilot. Under automatic control the flare may be a nearly constant-attitude maneuver, with sink rate reduced by direct lift control. The same aircraft under manual control may require rotation to flare. Such lack of harmony between the aircraft motions in manual and automatic operation makes the pilot's monitoring more difficult. Although it is known to be a significant factor in pilot acceptance, we currently do not have a good quantitative appreciation for motion harmony requirements, so this factor will remain qualitative until further research is conducted.

A final important criterion for pilot acceptance is the pilot workload required to perform an approach and landing. Several pilot workload measures and testing techniques are discussed subsequently in Subtopics D and E.

This completes our introduction to this section. We shall now consider the subject of psychomotor behavioral measurement in more detail.

## A. PSYCHOMOTOR BEHAVIORAL IDENTIFICATION TECHNIQUES

The dynamic response characteristics of human operators are important in a wide range of vehicular control situations. Psychologists and engineers have been studying specific tracking control situations for years and have found the human operator to be highly adaptable to a wide variety of machine dynamics. The use of systems analysis techniques together with dynamic response models of the human operator has tended to coalesce much of the apparently diverse and irreconcilable data, and provided a valuable construct for both system design and analysis. The dynamic response models of the human operator used in systems analysis activities have also proved to be extremely useful guides in designing man-machine experiments and defining relevant measurements of dynamic response performance (see Ref. 12).

The value of understanding pilot psychomotor behavior lies in the ability to predict results for a variety of conditions rather than relying on the demonstrated performance for a single set of conditions. This comes about as a result of defining the pilot's overall input-output behavior rather than just the explicit output performance.

With regard to human error, knowledge and specification of nominal behavior provides a basis for quantifying departures from such behavior, i.e., errors. For example, Ref. 13 reports the detection of a head-up display flight director tracking mistake as a result of monitoring a running estimate of the pilot's flight director-to-column transfer function. The pilot, following a minor distraction, began tracking the wrong symbol in the display. This was only a momentary error, and the pilot detected it himself*. But the incident did register clearly in the psychomotor behavioral measurements.

Another motivation for psychomotor behavioral identification is the simple fact that its frequent alternative, pilot performance measurement,

---

* The measurements also detected other types of errors committed by (but not mentioned by) some of the pilots. We shall illustrate and discuss the errors detected subsequently.

is often ambiguous. A given measured level of pilot performance can correspond to various combinations of:

- Pilot workload
- Controlled element response
- External disturbance level
- Displayed information.

In a general sense, to define psychomotor behavior is to define the input-output transfer relationships between all vehicle states (in the various ways they are perceived) and the various vehicle controls. This can be a formidable procedure, but there are some reasonable, feasible approaches.

The prime difficulty in attempting to measure psychomotor behavior is the understandable reluctance to hypothesize a behavioral model which then must be quantified experimentally. It is far easier and far less risky simply to measure and report the resultant pilot-vehicle performance, e.g., tracking error statistics. Quantification of behavior requires the experimenter to know what are the significant stimuli, the ways in which controls are functions of the stimuli, signficant noise sources, and the accompanying role of vehicle dynamics.

A complete survey of psychomotor behavioral measurement techniques and methods is well beyond the scope of this study. Much has been published under the heading of human operator identification and far more under the general heading of system identification. A few survey documents include Refs. 14-18. All we shall attempt to do is to outline a useful taxonomy of measurement approaches and to discuss what features are important to the measurement of human error. This effort is based on review of psychomotor measurements and techniques reported by many sources.

Finally, it is important to understand that the vast body of literature on this topic deals with single-loop control structure. Relatively few measurements have been made in a task-related multiloop context.

But a multiloop context is highly relevant to the consideration of human error, and we shall describe how to cope with this context in the next topic.

## B. APPROACHES TO PSYCHOMOTOR BEHAVIOR MEASUREMENT

In discussing the subject of psychomotor behavioral identification, it is first helpful to consider, in a general way, the important features of the identification process. A diagram of the general psychomotor identification process (or almost any identification process for that matter) is shown in Fig. 5. The central features are (i) the subject, (ii) the subject's stimuli and responses, and (iii) the model structure which reflects the psychomotor characteristics of the subject. The other features shown aid in producing a definition and quantification of the model structure and include the disturbance input, identification method, solution criteria, and search procedure. The interpretation of results is the means of conveying essential information to the experimenter. Each of these aspects will be discussed and followed by a discussion of what measurement features are most appropriate for simulator studies of human error.

### 1. A Taxonomy of Psychomotor Measurement

a. Model Structure. An important step in defining a psychomotor behavior measurement approach for multiloop tasks is the choice of model structure. Without a definable, explicit model structure, there is no real basis for quantification of the stimulus-response functional relationships. The reticence of some investigators to hypothesize a model structure has blunted the interpretation and usefulness of many experimental results.

The model structure is simply the framework about which measurements can be quantified and organized. This framework can take on many forms and degrees of complexity, however.

In general some kind of parametrical expression is needed in order to interpret, summarize, and compare results. However, parametrical

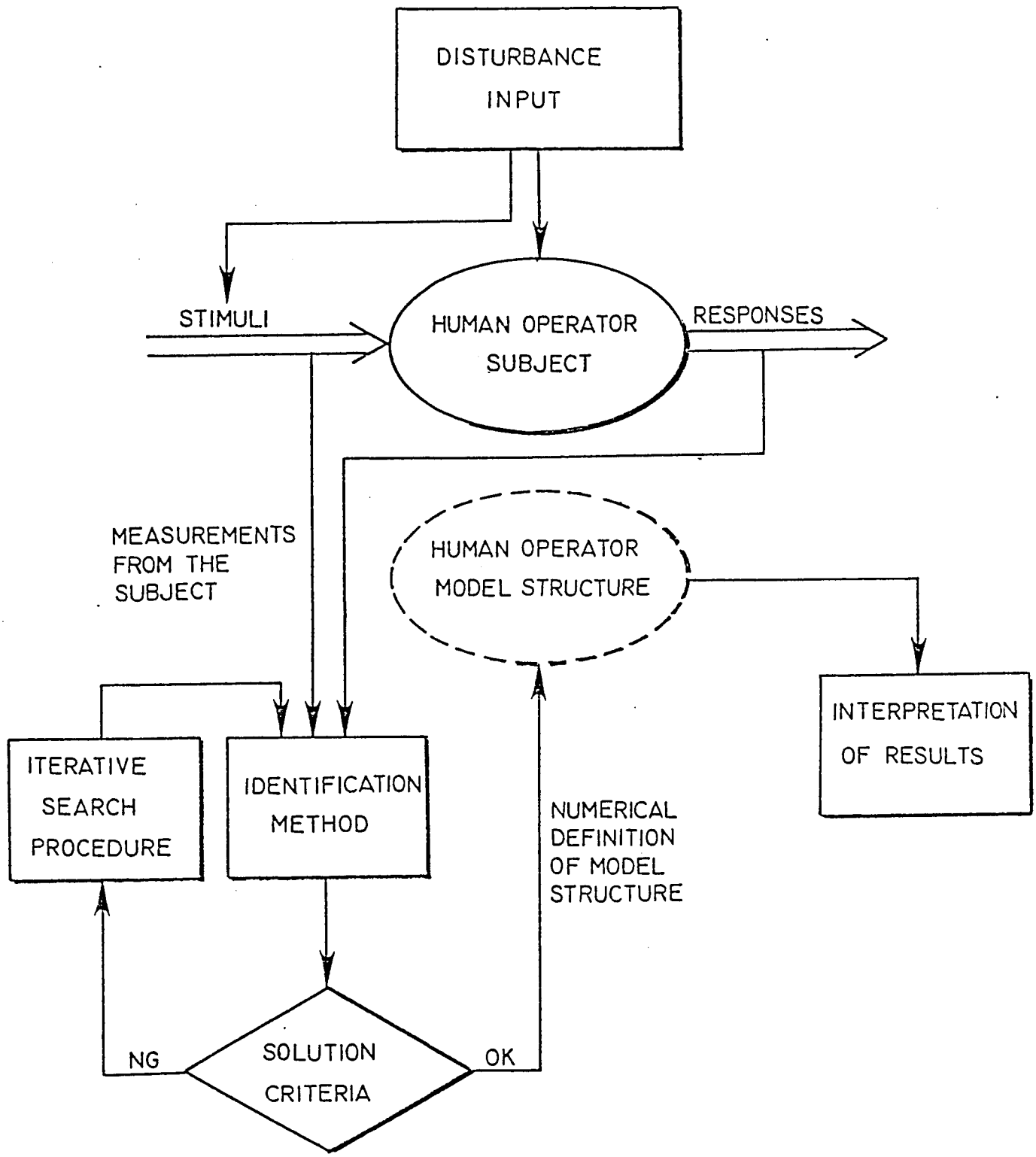



Figure 5. The Identification Process

features can be expressed after having identified behavior in a non-parametric form. For example, the results of using spectral analysis techniques to obtain human operator describing functions are expressed in a general frequency domain form (amplitude and phase) without reference to specific pilot behavioral parameters. Such results can then be interpreted in terms of summary parameters as stimulus-response amplitude and phase at a few specific frequencies of interest.

The exact nature of parameters chosen to represent human operator behavior is an important issue. In the role of safety verification analysis, peak excursions or standard deviations in, say, altitude or airspeed are obviously important parameters in judging terrain clearance or stall safety margins, respectively. As for human error evident in psychomotor behavior, it was suggested in Ref. 1 that effective system bandwidth for a particular task is a fundamental parameter*. It can, for example, help to establish the level of successive organization of perception at which the pilot is accomplishing that task. Changes in a time history of effective bandwidth can also serve as both event markers and error indicators.

Other parameters which reveal pilot behavior in direct ways are:

- Pilot's stimulus-response phase angle at or near the effective pilot-vehicle system bandwidth (this is an indication of lead compensation or anticipation and therefore workload).
- Pilot's stimulus-response gain at or near the effective system bandwidth (if evaluated as a function of time or specific events this can be an indication of appropriate adjustment to changing conditions).
- Non-zero crossfeeds or feedforward actions which coordinate controls in various special tasks (this can be a direct indication of pursuit or precognitive behavior as opposed to strictly compensatory).

---

* Gain crossover frequency is a convenient measure of system bandwidth.

Perhaps the most fundamental division in model structure is between structural isomorphic and algorithmic models (Ref. 19).

An isomorphic structure refers to having a form much like that of the human operator or the operator's organizational structure. Isomorphic can apply to neuromuscular, sensory, and equalization functions such as shown in Fig. 6 (from Ref. 19). Taken on a larger scale, it can also apply to a basic task-dependent loop structure as demonstrated in Figs. 7 and 8 for two common aircraft maneuvers.

The algorithmic psychomotor behavior structure supplies (e.g., Fig. 9) the various organizational units which are, in turn, identified or measured by any suitable identification method — parametric or nonparametric, time or frequency domain.

An algorithmic model structure is, in some ways, an abstraction of psychomotor behavior and is based on the notions of optimal control-optimal estimation, i.e., modern control theory. Typically this form of model expresses the human operator's adaptive control (motor) behavior as an optimal controller which makes use of all system states and controls in such a way as to minimize some form of cost function. Those state variables which are assumed to be perceived are operated on by an optimal estimation process (Kalman filter) in order to generate the needed states for the control process. Much success has been achieved with this approach as illustrated in Refs. 20, 21, and 22.

Three areas of difficulty of the modern control theory algorithmic model approach regarding psychomotor behavior are given in Ref. 19. These are, briefly stated,

- The human operator must contain essentially complete knowledge of the man-machine characteristics, i.e., be a complete internal model. Although this might be plausible at the precognitive level of skill development, it is incompatible with what we know about the compensatory level.
- Identification from experimenal data is difficult.

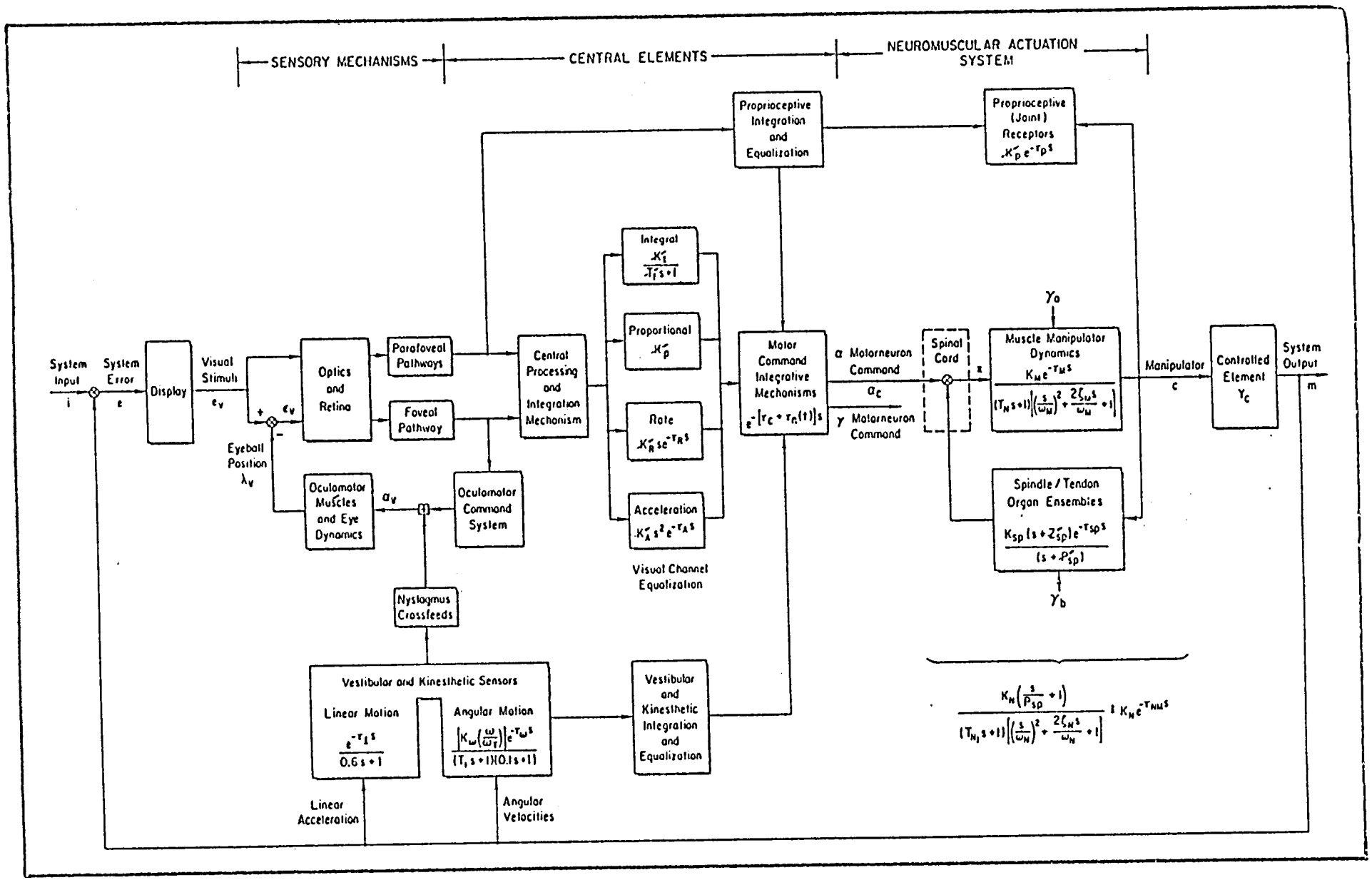



Fig. 6. Structural isomorphic model of man-machine system.

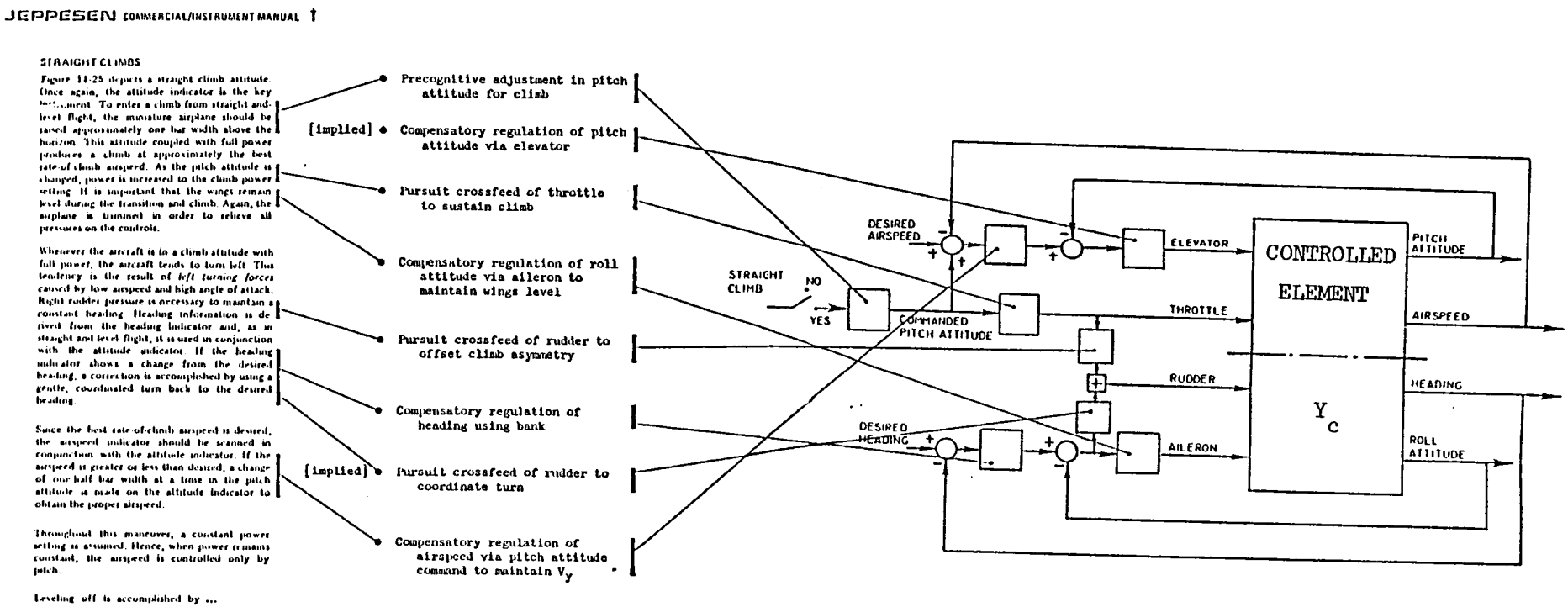



Figure 7

Translation of a Verbal Task Description to a Pilot-Vehicle Loop Structure
(Straight Climb on Instruments)

† JEPPESEN Commercial Instrument Manual, Jeppesen Sanderson Inc., Denver, CO, 1977.

BASIC HELICOPTER HANDBOOK AC 61-13
pp. 81-82

... 2. Initiate the approach by lowering the collective pitch control the amount required to descend at an angle of approximately 12° on the final approach leg. As collective pitch is lowered, increase right pedal pressure as necessary to compensate for the change in torque reaction to maintain heading and adjust throttle to maintain proper RPM. Maintain attitude with cyclic control until airspeed nears approach airspeed then adjust to the attitude (with cyclic) that will maintain this approach airspeed.

3. The angle of descent and rate of descent are primarily controlled by collective pitch, the airspeed is primarily controlled by the cyclic control, and heading on final approach is maintained with pedal control. However, only by the coordination of all controls can the approach be accomplished successfully.

4. The approach airspeed should be maintained until the point on the approach is reached where, through evaluation of apparent groundspeed, it is determined that forward airspeed must be progressively decreased in order to arrive at hovering altitude and attitude at the intended landing spot with zero groundspeed.

5. As forward airspeed is gradually reduced by the application of rearward cyclic pressure, additional power (collective pitch) must be applied to compensate for the decrease in translational lift and to maintain the proper angle of descent. As collective pitch is increased, left pedal pressure must be increased to maintain heading, throttle adjusted to maintain RPM, and cyclic pitch coordinated to maintain the proper change in forward airspeed.

6. The approach is terminated at hovering altitude above the intended landing point with zero groundspeed. Very little, if any, additional power should be required to stop the forward movement and rate of descent of the helicopter if power has been properly applied during the final portion of the approach.

7. If the condition of the landing spot is unknown, the approach may be terminated just short of the spot so that it can be checked before moving forward for the landing. At altitude, a ship should be used to align the fuselage with the ground track. The rotor is tilted into the wind (with cyclic pressure) enough so that the sideward movement of the helicopter and wind drift counteract each other. Heading is maintained along the ground track with the antitorque pedals. (See "Crosswind Consideration During Takeoffs.") This technique should be used on any type of crosswind approach - shallow, normal, or steep.

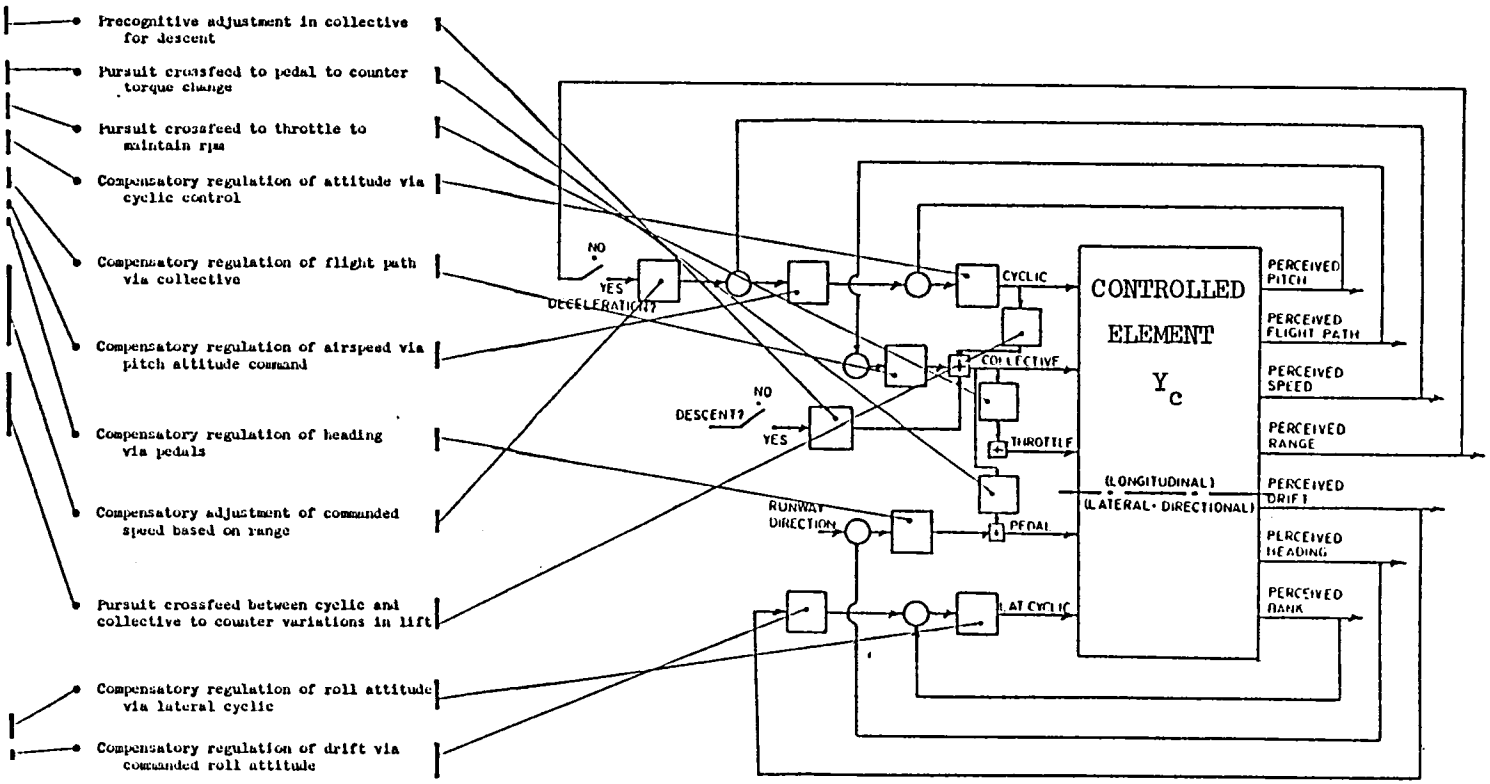



Figure 8

Translation of a Verbal Task Description to a Pilot-Vehicle Loop Structure
(Helicopter Approach to Hover)

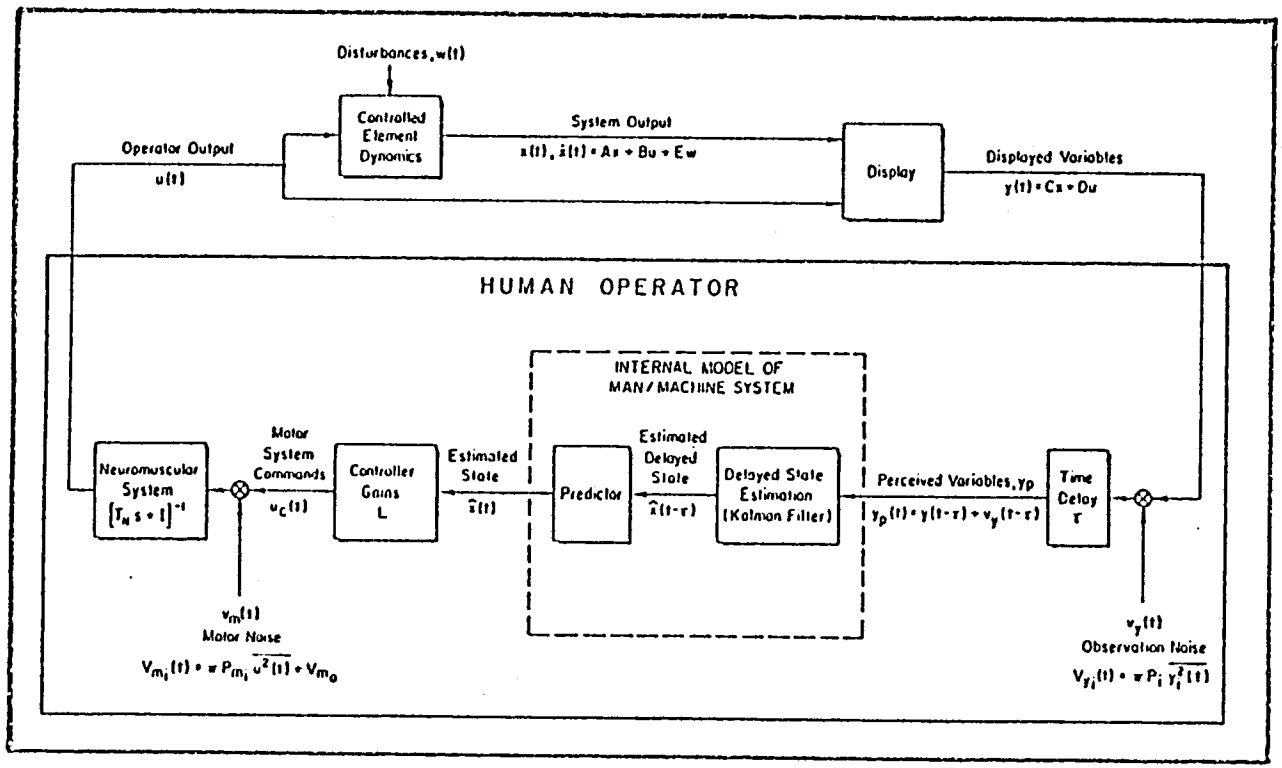



Fig. 9. Algorithmic (linear optimal control) model of man-machine system.

- A cost function appropriate to a particular task must be available.

b. Identification Method. The identification method consists of the computational manipulation of basic stimulus and response data in order to quantify the model structure. The following general methods have been applied to human operator behavior measurements:

| | | |
|---|---|---|
| Frequency Domain | Fourier analysis by FFT* | Non-parametric |
| | Cross-spectral analysis | |
| Time Domain | Variance analysis | |
| | Cross-correlation analysis | |
| | Response error analysis | Parametric |
| | Equation error analysis | |

As indicated, two broad kinds of classifications are (i) time versus frequency domain and (ii) non-parametric versus parametric.

Fourier analysis has been widely used for measuring psychomotor behavior (e.g., Refs. 23 and 24). The attractiveness of Fourier methods stems from the capability they provide for making on-line FFT measurements with high signal-to-noise ratio. The resulting describing functions, error variance, relative coherence, and remnant are usually computed off-line. The method requires the use of prescribed sums of sinusoidal inputs to the pilot (error signal) and a known controlled element to compute finite Fourier transforms. The result is a spectral description of the pilot's response to a particular signal at several discrete frequencies. These data are then frequently fitted by an effective model structure in order to obtain specific values of effective neuromuscular delay, equalization, and remnant.

Cross-spectral analysis requires only that various spectral and cross-spectral density functions of the pilot's input and output signals be measured (Ref. 12). The cross spectra are computed with respect to a common signal, and thus the pilot's input-to-output describing function

---

* FFT = Finite Fourier Transform.

is derived from the ratio of the input-to-output cross spectra so computed. References 25 and 26 provide good examples of applying the cross-spectral technique. One of the advantages of both Fourier analysis and cross-spectral techniques is that they provide remnant spectral density as well as describing functions. One of the drawbacks to both the Fourier analysis and cross spectral methods, however, is that relatively long run lengths in time are required to get good, low frequency data for describing functions. A discussion of this problem can be found in Ref. 27.

Variance analysis is the most extensively used time domain method, but its value is limited. The most direct application is safety verification analysis, i.e., estimating the probability of exceedence of nominal limits such as with airspeed or flight path. Variances, per se, are not highly sensitive to changes in workload or psychomotor behavior. Although seldom used, some estimation of effective bandwidth can be obtained from one-half the null crossing frequency or from the ratio of rate variance to displacement variance, i.e.,

$$\omega^2_{eff} \doteq \frac{\sigma^2_{RATE}}{\sigma^2_{DISP}} \qquad \text{(Ref. 28)}$$

"Bandwidth" is a vague term unless the signal spectrum is rectangular. For other spectral shapes the dimensionless variability can be used to define a rectangular bandwidth equivalent, i.e.,

$$\omega_{i_e} = \frac{\left[\int_o^\infty \Phi_{xx}(\omega)\, d\omega\right]^2}{\int_o^\infty \left[\Phi_{xx}(\omega)\right]^2 d\omega} \qquad \text{(ref. 29)}$$

This is the bandwidth of a hypothetical rectangular filter which would pass a signal x with the same mean squared statistical error as the actual filter when the input is white noise. $\Phi_{xx}(\omega)$ is the power spectral density of signal x, where the signal variance $\sigma^2_x$ is defined as

$$\sigma_x^2 = \int_0^\infty \Phi_{xx}(\omega)\, d\omega$$

Cross correlation analysis is a time domain method which has been used to describe non-linear, non-stationary weighting functions of a human operator (Refs. 30 and 31).

Response error and equation error methods, also known as parameter trackers, have enjoyed much popularity in identifying inanimate systems and also appear useful in measuring human operator behavior. One of the key advantages of the response error and equation error methods, especially for identifying human error, is the aspect of revealing fairly abrupt changes with respect to a time-line of events by means of a short term averaging technique employing a sliding window. Reference 18 treats these methods in a general way, but there are many variations (e.g., Refs. 32 and 33). One has the freedom to adapt these methods to specific characterizations (model structures) of the psychomotor behavior. More will be said about this subsequently.

c. Command and Disturbance Inputs. Excitation of the pilot-vehicle system is essential for any type of psychomotor measurement. As mentioned previously, the human operator must be induced to interact with a simulation — to follow commands, to regulate against disturbances by closing loops or otherwise to perform required tasks. Commonly used command and disturbance inputs include both deterministic and random signals listed in Table 5.

Some human operator identification schemes include a disturbance input which can be adapted as an integral part of the scheme. In the case of the describing function analyzer (DFA) (Refs. 34, 35, and 36), a sum of sine waves is provided. It can be employed either as a disturbance or as a command, and the operator's describing function at the sinusoidal frequencies can be computed quite directly from knowledge of the resulting control movement. In addition the remnant can be computed by the serial segments method (Ref. 37).

TABLE 5

EXAMPLES OF SIMULATED INPUTS SUITABLE FOR PSYCHOMOTOR MEASUREMENTS

A. Random or quasi-random signals (unpredictable by definition) necessary for identifying compensatory level of skill development

   1. Representing forms of

      a. Atmospheric turbulence

      b. Radio guidance anomalies

   2. Alternative generating sources

      a. Continuous or discrete Gaussian stochastic signal sources

      b. Quasi-random sums of five or more sine waves

B. Deterministic, but unpredictable signals — necessary for identifying possible transitions to levels of skill development higher than compensatory

   1. Representing forms of

      a. Discrete gusts

      b. Wind shear

      c. Radio guidance anomalies

      d. Intrusions which lead to evasive action by pull-up or side-step maneuvers

      e. Engine failures which lead to abrupt moments and forces on the aircraft

      f. Cockpit warning and caution signals

      g. ATC commands, advisories, responses

   2. Alternative generating sources

      a. Transient signals, e.g., steps, pulses, ramps, versines

      b. Pseudo-random binary signals

      c. Voice commands, responses

C. Deterministic and quasi-predictable signals — necessary for identifying pursuit and precognitive levels of skill development

1. Representing forms of

a. Marker beacon signals, to/from signals, event markers

b. Checklists

c. ATC commands, advisories, responses

d. Familiar features of terrain, especially on visual approach routes

e. Moving maps and elevation profiles of routes

f. PPI of relative motions of neighboring traffic and weather

g. Pilot-induced oscillations

h. Low frequency, lightly damped vehicle modes

i. Optical landing guidance anomalies caused by ship motions in a coherent sea

2. Alternative generating sources

a. Single sine wave

b. Sums of a few sine waves

c. Narrow-band processes, in general

d. Oscillators

e. Event markers

f. Voice commands, responses

D. Deterministic signals which are useful as injected test inputs for identifying inanimate systems such as controlled elements

1. Representing forms of

Typical control and disturbance inputs

2. Alternative generating sources

a. Sum of sine waves (Ref. 23 and 34)

b. Frequency sweep (Ref. 38,)

c. Pseudo-random binary (Refs. 39 and 40)

Other measurement approaches may take advantage of disturbance inputs provided within the simulation, such as radio guidance anomalies, atmospheric turbulence, or wind shears. Whatever the generating source, one must be careful not to compromise realism (and thereby to compromise pilot motivation) when adapting the disturbance input to provide adequate signal-to-noise ratio for the purpose of identifying describing functions and remnant over the desired measurement bandwidth. Sums of sine waves, in general, provide the superior signal-to-noise ratio essential for identifying remnant. Further discussion of the various sources of remnant can be found in Ref. 12.

Some identification techniques will identify the inverse plant instead of the human operator when significant amounts of remnant are present (Ref. 41 demonstrates this phenomenon). The parameter model identification scheme, however, will still accurately identify the pilot even when large amounts of pilot remnant are present. This unique feature, along with other attributes of the parameter model identification scheme, are demonstrated in Ref. 13 where it is applied to a realtime, piloted simulation. Some selected results from Ref. 13 are contained in the next subsection.

Further discussion of the identification of elements within a closed loop can be found in Refs. 12, 41, and 42.

d. Solution Criteria. For any particular psychomotor behavioral measurement approach it is necessary to judge how well the identification method has produced quantification of the model structure. According to Ref. 43, solution criteria can be classified as:

Error minimization

Likelihood approach

Prediction error

F-ratio

These include the popular least squares and maximum likelihood criteria. The least-squares method is perhaps the most commonly used parametric identification method; see Ref. 14. Among its advantages are that it is

easy to apply, quick to use, and the calculations can easily be performed recursively in the observed data. These advantages permit real-time, on-line identification in a simulation environment (e.g., where the required data is in a high speed digital computer). The chief disadvantage of the least-squares method is that it does not permit modeling of the noise structure of the system, and that it gives biased estimates unless the noise structure is of a certain type. These problems have not been found troublesome, however, in the psychomotor measures of Refs. 13 and 44.

The maximum likelihood method has been widely used in all types of system identification (see Ref. 16 for a partial list). The major limitation in connection with simulation is the need for considerable computational power. The advantage of the maximum likelihood method is unbiased estimates. However, it is more difficult to apply than the least squares method and requires much more computational power.

e. Iterative Search Procedure. In some cases it is necessary to apply a search procedure in order to converge upon a solution to a given identification method. This is a technical matter which is of little concern here except to note its role. In many identification approaches, a direct solution to model structure is possible, and a search procedure is unnecessary.

Some of the search procedures which are available include:

Manual

PARTAN

Davidon-Fletcher-Powell-Levenberg

Newton-Raphson

Random

Simplex

A discussion of specific search procedures is beyond the scope of this report.

f. Presentation and Interpretation of Results. In order to appreciate the results of psychomotor behavior measurements, the experimenter may need to examine more than just the numerical definition of whatever model structure is employed. For example, if the psychomotor model is in the form of several finite difference equation coefficients (i.e., time domain), then it may be useful to display an indication of frequency domain quantities such as effective bandwidth or phase shift at a particular frequency of interest. (In fact, Ref. 44 demonstrates that the "raw" difference equation coefficients can behave very strangely under certain circumstances, but that frequency domain parameters are very well behaved.) Or, as an example of the converse, a non-parametric cross-spectral measurement might be better summarized in terms of an effective neuro-muscular delay or lead time constant.

The point is that any basic behavioral identification scheme can be further manipulated to provide indications convenient to the experimenter. A particular model structure and identification method may be efficient thus permitting realtime computation and data reduction, but subsequent transformation to different terms may be of more direct benefit.

2. Measurement Approaches Appropriate To Human Error

a. Diverted Attention. Diverted attention from flying the aircraft and spontaneous improper actions are believed to be sources of human error underlying many of the cited causes in Ref. 2 which involve a flying error. Measurement techniques are highly developed for identifying the role of diverted attention from flying the aircraft as a source of human error, provided the flying tasks for each phase of the mission have been carefully defined at the outset of an experiment. The most prominent effects of diverted or divided attention are to reduce the pilot gain and to increase remnant in the affected channels of attention for which psychomotor measurement methods have already been discussed.

b. Spontaneous Improper Action. Measurement techniques are also well-developed for identifying spontaneous improper actions, provided

the sequences of tasks and actions necessary for mission success and failure have been thoroughly planned and defined at the outset of an experiment. Such careful pre-experimental identification of procedures, both proper and improper, provides a framework exemplifying the spatial-temporal facets of the mission phase event- or time-line which are essential to the recognition and interpretation of "slips" at the precognitive level of operational behavior. The necessarily thorough pre-experimental definition of procedures was applied in Refs. 4 and 5, but in Ref. 5 the details of recording discrete actions such as setting switches or levers, responding to checklists, or coping with emergencies were relegated to an observer's commenting on a voice recorder, coupled with voice records of all flight deck communications. Since retrieval of "slips" from voice records is both tedious and cumbersome, as well as subject to the additional interpretation of the observer and participant, it is preferable to institute automatic recording of discrete actions by the crew members wherever possible. Thereafter to detect "slips" it is possible to employ automatic comparison of the recorded time-line of discrete actions with the pre-experimentally recorded time-line of "normal" and "emergency" procedures established for the scenario.

c. Monitoring and Decision-Making Errors. With increased use of automatic controls and computers in modern day aircraft and traffic control systems, the role of the human operator is becoming more supervisory, involving increased amounts of monitoring and decision making. In these roles, human outputs are typically discrete (as opposed to continuous control actions) and include non-manual actions such as verbal communication. Monitoring and decision making errors can arise due to misperception of monitored information and misinterpretation of perceived information. Errors can also occur in the more cognitive aspects of decision making where the operator must account for various possible consequences of the alternative actions available to him. Again, since retrieval of monitoring and decision-making errors from voice records is tedious and cumbersome, it is possible to employ automatic comparison of the recorded time history of discrete actions with the pre-experimentally recorded time-line of normal and emergency procedures established beforehand for the scenario.

Monitoring and decision-making constructs and viewpoints are useful in full mission simulations with a complete crew in several ways. First, human errors sometimes appear to be inexplicable when, for example, only two courses of action are possible, and an operator appears to make the obviously wrong choice. By considering the elements of these task situations in a decision-making context one can gain additional insight into the underlying factors involved. Second, if specific analytic decision-making models are reasonably appropriate descriptors of the mission phases being simulated, then the model can serve as a means for the analysis and interpretation of the experimental results. Third, a combination of monitoring, decision-making, and control viewpoints is essential in treating repeated simulation runs by one crew, or an ensemble of simulations involving many crews. In a single run, behavior and performance for all the tasks involve specific concrete actions (or inactions) flowing in a sequence. Error is identified as an extreme deviation from a desired state. With many runs, these concrete actions often exhibit differences, either in kind or in degree. A probabilistic structure for particular events then becomes appropriate as a means of describing the experimental data. Further, the potential tradeoffs (based on experience and training) involved in selecting various emergency actions can be exposed in the light of a utility concept. Monitoring and decision making theories are appropriate for such considerations.

For simulations where a monitoring and decision making construct is likely to be useful, the experimenter must recognize this potential at the outset by appropriately structuring the experimental tasks, scenarios, and performance measures. Then, when particular models for decision making are to be considered in data analysis, there may be further impact on the experimental design.

In Ref. 1, monitoring and decision making are first presented from a conceptual point of view in order to identify the basic components of monitoring and decision making tasks that must be taken into account in simulation setup, selection of measurements, and experimental design. Analytical procedures for data analysis and modeling are then briefly

covered. In the most general approach to studying monitoring and decision-making behavior, the detailed structure of the operator's task may not be clear so that only very general data analysis procedures can be applied with any certainty. As more is understood about the operator's behavior, certain assumptions may be invoked to allow more detailed analysis and perhaps modeling of the operator's task. Reference 1 concludes with an example to illustrate how a specific situation can be analyzed from a decision perspective to discover factors important in developing the appropriate experimental measurements to be made in a simulation.

d. An Example Identifying Control Task Errors. Using the measurement taxonomy outlined in the previous subsection, the following approaches are recommended:

Model structure — keep it as simple as possible while observing all significant features within the nominal piloting task. It may be necessary to make successive refinements, each more complex, in order to settle on an optimum model structure.

Identification method — time domain analysis may be more sensitive to revealing human error events than frequency domain analysis. One successful direct method using a specific isomorphic model structure is the least squares (equation error) parametric method described in Refs. 13, 41, 44, and 45.

Disturbance inputs — existing atmospheric turbulence is capable of providing the needed disturbance but must be strong enough to predominate over pilot remnant.

Solution criteria — least squares fitting using accumulated data is adequate. Non-stationary effects may be obtained by restarting identification periodically or by dropping off old data as new data are acquired (sliding window concept, Ref. 13; or fading memory, Ref. 32).

Search procedure — none is required for a least squares parameter method, per se, but it may be useful to carry along more than one model structure or identification method and to search for the best solution according to goodness of fit.

Interpretation of results — make use of results in realtime if possible. Notify the experimenter about anomalies as soon as detected to signal possible human error events. Attempt to correlate subjective and objective, e.g., performance with effective bandwidth, workload with phase angle shifts, successive organization of perception with appearance of feedforward, or cross-feed paths as well as with effective bandwidth.

The time histories shown in Figs. 10 and 11 demonstrate how a pilot control strategy identification scheme can be used to identify and quantify human error. The time histories were taken from a realtime, piloted simulation and represent a pilot controlling a conventional jet transport aircraft on final approach. In Fig. 10 the pilot was using a standard head-down flight director, and in Fig. 11 the pilot was using a flight path head-up display (HUD). The non-intrusive pilot identification program (NIPIP) described in Ref. 13 was used to measure the pilot's control strategy (labeled as $\hat{Y}_p$ ($j\omega$) in Figs. 10 and 11) as well as the bandwidth and phase margin of the combined pilot-vehicle system (labeled as $\omega_{c_{FD}}$ and $\phi_{m_{FD}}$ in Figs. 10 and 11). The bandwidth (which is also called the crossover frequency) reflects how tightly the vehicle is being controlled. Higher bandwidths are desirable because the combined pilot-vehicle system is less responsive to external disturbances. To achieve higher bandwidths, however, requires higher workload by the pilot. The phase margin reflects the relative stability of the combined pilot-vehicle system (positive, zero, and negative values of $\phi_m$ correspond to stable, neutrally stable, and unstable systems, respectively). Reference 13 reports that the phase margin was particularly sensitive to changes in pilot control strategy and could be used to identify and quantify certain types of pilot error (specifically, errors in control strategy). Examples of this are shown in both Figs. 10 and 11).

In Fig. 10 the pilot makes a "control reversal," which is labeled as Item 6 in the figure. That is, the pilot put in a pitch up command when the rate and position of the flight director called for a pitch down command. In Fig. 11 the pilot started tracking the wrong symbol in the HUD (specifically, the glide slope symbol instead of the flight path symbol), which caused the flight director to diverge. After a few

① Outer Marker
② Time Window Full
③ Middle Marker
④ Breakout (i.e., below cloud cover)
⑤ Touch Down
⑥ Control Reversal
⑦ Pilot Flying "Open-Loop"
⑧ Note the Sudden Change in Phase Margin

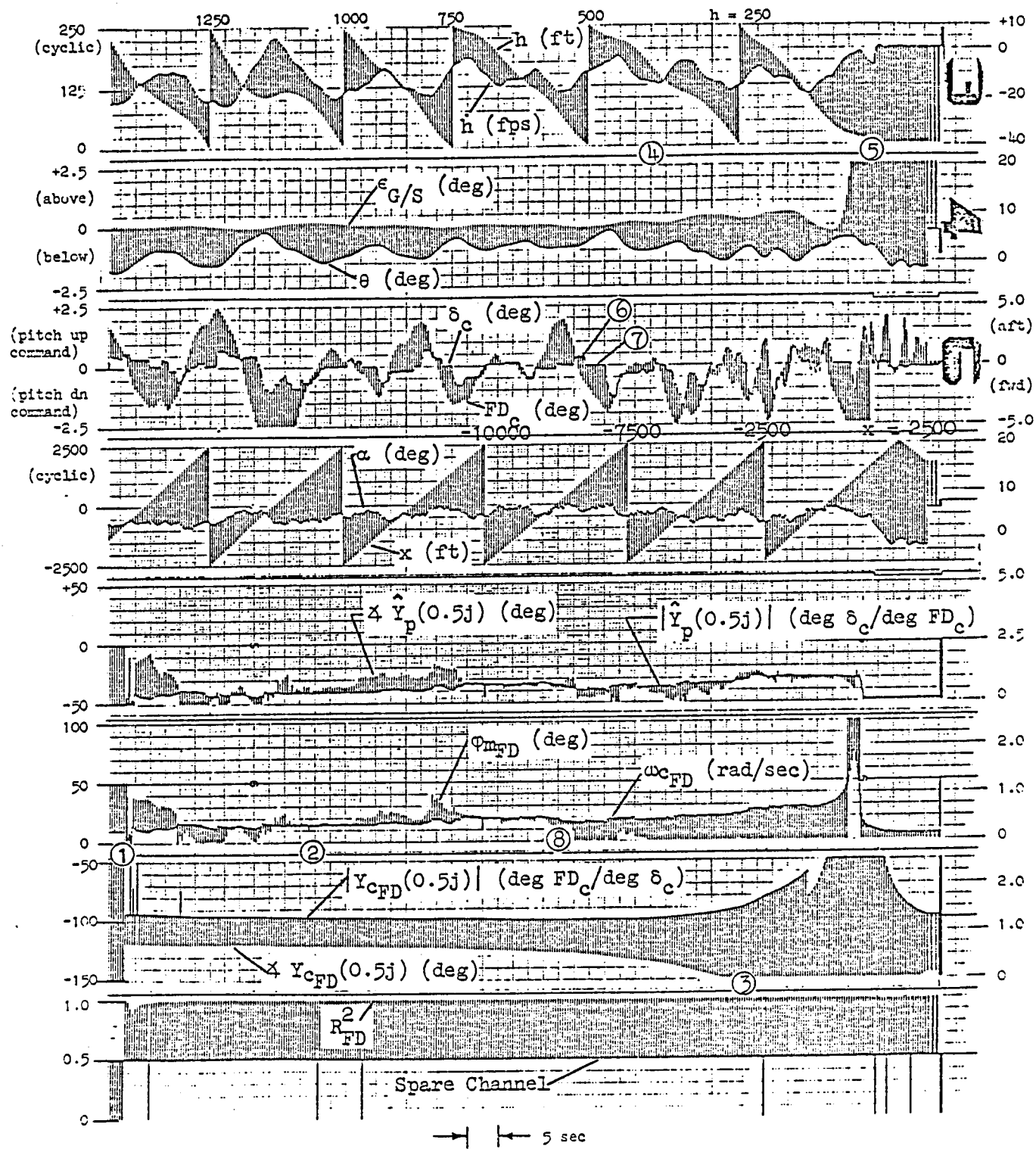



Figure 10. Time History on Final Approach with Head-Down Flight Director

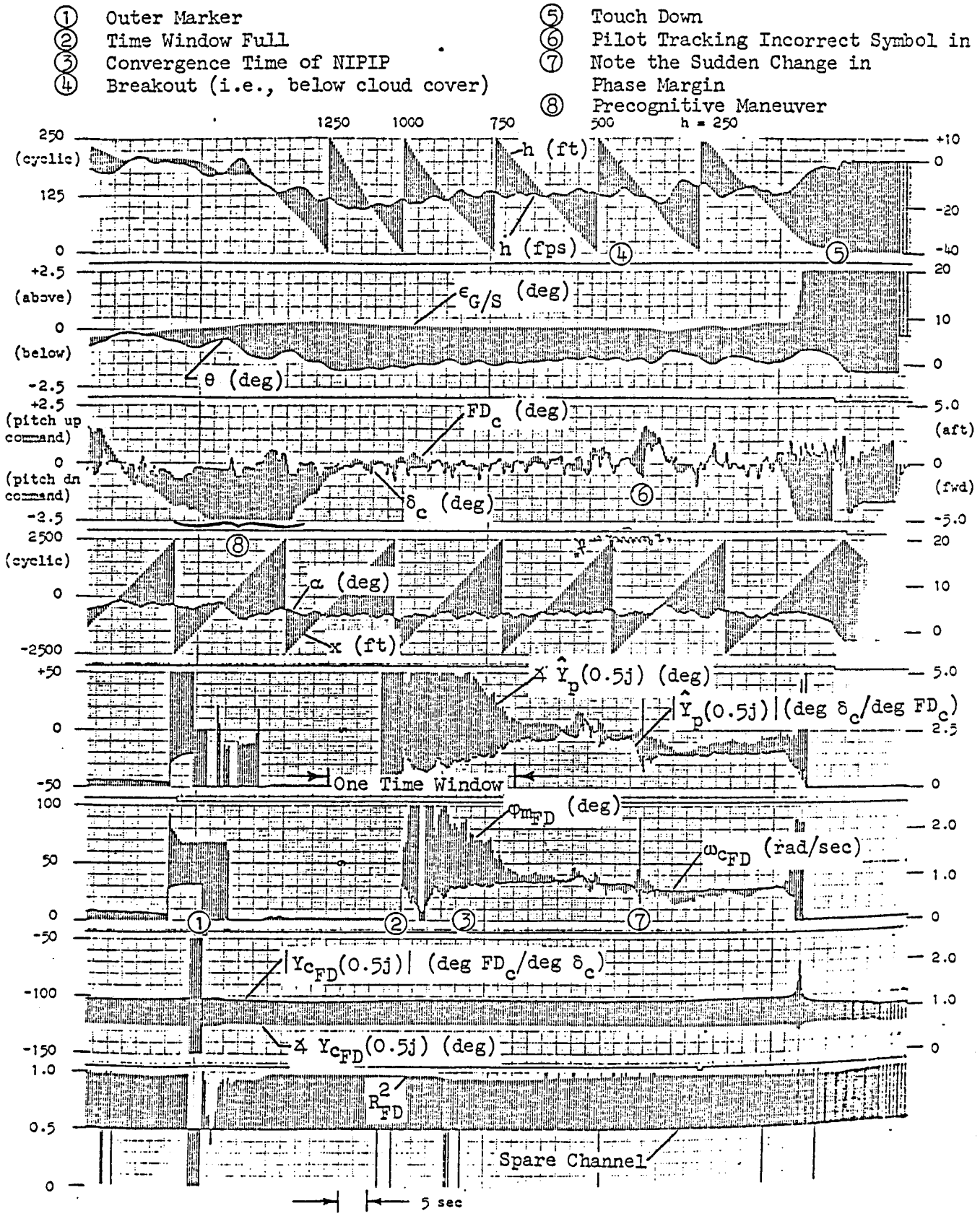



Figure 11. Time History on Final Approach
Using Flight Path HUD with Flight Director

seconds, the pilot realized his error and made a large corrective control input. Both errors were quantified by large and sudden changes in $\hat{Y}_p$ $(j\omega)$ and $\phi_m$. Changes in the bandwidth and/or phase margin may also reflect other events such as pilot distraction or changes in pilot workload. Learning effects and skill retention are also quantifiable with bandwidth and phase margin.

## C. MEASURING HUMAN RESPONSE TO A CHANGE IN THE (CONTROL) TASK SITUATION

Some of the most critical events in the context of both flight control and air traffic control will involve changes in the task situation or organization of pilot activity, particularly in failure management and other emergency situations. Critical control events typically involve a change in the task situation or organization of activity. This could consist of a planned event such as entering a terminal area and following a Standard Terminal Arrival Route (STAR), or an unexpected event such as a system failure or system deviation. On the one hand, the required pilot activity could consist of:

- Changes in the organization of manual control activity from compensatory to precognitive and back such as executing a side-step maneuver on final approach to parallel runways or pushing over to intercept the glide slope.
- Changes in the organization of manual control activity from compensatory to pursuit and back such as executing a Standard Instrument Departure (SID) or STAR with the aid of a moving map display.
- Manual control action such as taking over from an automatic system and continuing to fly the vehicle manually at a pursuit (rather than compensatory) level in the organization of perception.
- Monitoring and decision response such as switching to backup system from a primary system in response to a warning indicator or other displayed indication of emergency or failure.

On the other hand, the required traffic controller activity could consist of:

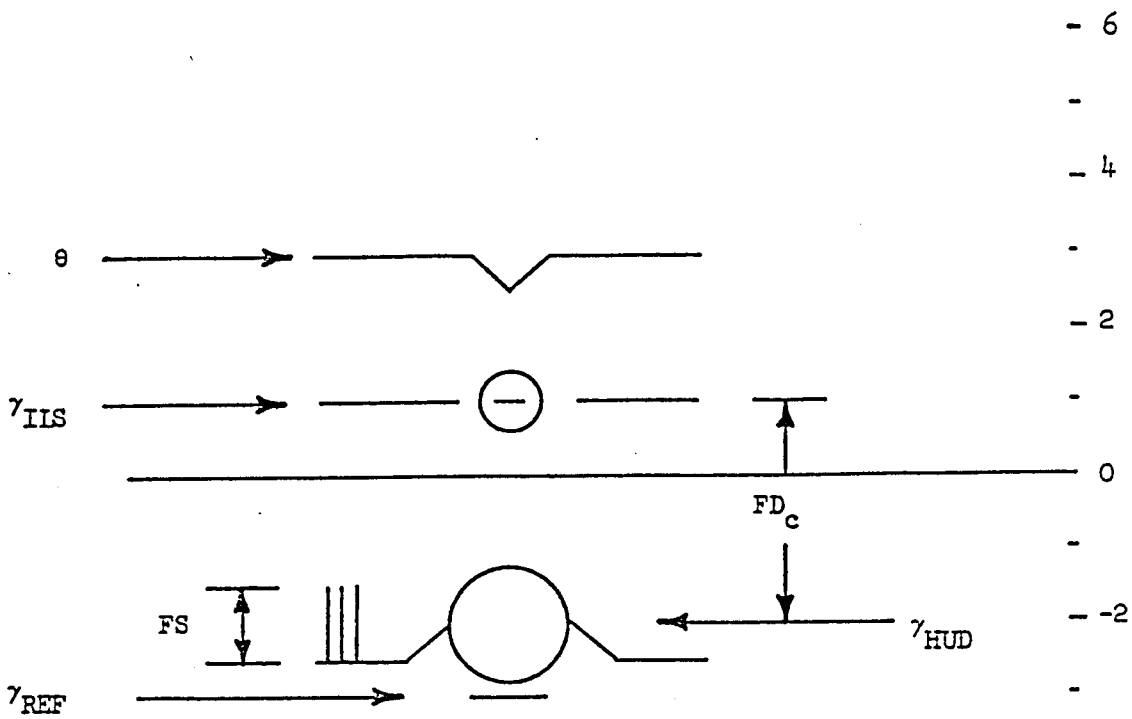
- Issuing a procedural advisory about potentially conflicting traffic
- Commanding evasive action.

In each case, task performance is strongly affected by the degree of expectancy and level of training and practice. For both types of activity within flight or ground context, the reaction times could be comparable (for similar stimulus levels). However, the initial decision and discrete switching or advisory action may solve the problem in the procedural task case, while continuous subsequent activity is required in the control case or command case.

### 1. Change in the Organization of Manual Control Actions

A graphic demonstration of how a pilot changes his organization of perception is contained in Fig. 11. The pilot is controlling a CTOL aircraft on final approach with the aid of the head-up display depicted in Fig. 12. The pilot is initially flying straight and level, and he must transition to a descending three-degree flight path angle and capture the glide slope.

Note from Fig. 11 that by the time the pilot has reached the outer marker he has not yet performed the required transition. He has flown through the glide slope (viz., $\varepsilon_{G/S}$ in the figure) and now the flight director, $FD_c$, is commnding a large pitch down angle. The pilot does <u>not</u>, however, follow the flight director commands, as evidenced by the lack of activity in the control column, $\delta_c$. The pilot is probably performing a <u>precognitive</u> maneuver. He pulls back on the throttles (not shown in the figure) and pitches down in order to get the aircraft to descend, based on his knowledge of the aircraft dynamics. Thus, during the transition phase the pilot is monitoring or closing a very loose loop on flight path angle. When the aircraft gets close to the desired




Display control laws:

$$FD_c = \Delta\gamma_V + \frac{\ell_x}{V}\dot{\theta} + K_{\epsilon 2}\epsilon_{G/S} + \frac{0.4\,s}{s + 2.5}\,\theta$$

$$FS = \frac{1.0}{s + 1.0}\,(V - V_{ref})$$

$$\gamma_{HUD} = \gamma_V + \frac{\ell_x}{V}\dot{\theta} + \frac{0.4\,s}{s + 2.5}\,\theta$$

$$\gamma_{ILS} = -\gamma_{REF} - K_\epsilon \epsilon_{G/S}$$

$\gamma_{REF}$ = ILS glide slope (= 3.0 deg)

$\gamma_V$ = Inertial flight path angle of aircraft

Status: Aircraft is pitched up 3 deg and on a 2 deg flight path angle. Glide slope is 3 deg above aircraft.

Figure 12. Simplified Sketch of Flight Path HUD with Flight Director (Longitudinal Axis Only).

flight path angle the pilot reverts to compensatory tracking of the flight director with the control column.

Note from Fig. 11 that compensatory $\hat{Y}_p$ is virtually zero during the precognitive maneuver, but converges to a reasonable solution rapidly once the pilot starts tracking the flight director. This is because NIPIP was designed to measure the pilot dynamics in a compensatory tracking task only and not during precognitive maneuvers. It may be possible, however, to quantify this precognitive maneuver by examining a state space similar to the one used for modeling flare maneuvers (Ref. 43).

2. Change in the Controlled Element

Some research has been accomplished by STI (Refs. 47 and 48) and others (e.g., Refs. 49 and 50) in efforts to measure and interpret operator and system performance when there is a sudden change in the manually controlled element. Early work by Sadoff (Ref. 49) considered pilot control with pitch damper failures in a centrifuge simulator. These and other data were brought together by STI to obtain a model for interpreting the pilot's response to a task "transition" which contains four phases:

- Pre-transition steady state
- Post-transition "retention," where the pilot has not yet reacted properly to the transition
- Transition control, where the pilot may use large corrective control actions to stabilize the system and reduce large errors which may build up during retention
- Post-transition steady state

The two middle phases are the key to transition performance. With high expectancy (transition probability) the retention time is short; and this might be the case during approach, while failures of the flight control system (FCS) during en route phases would be unexpected and result in longer retention times.

For difficult dynamic transitions, training is particularly important. For example, in our studies (Ref. 48) skilled pilots completely lost control during the first 20 to 30 attempts to handle a severe control system failure, but after 200 trials their response was nearly time-optimal with very little perturbation in system error. The question now at hand is: How does lapse in practice affect this highly trained state, and what type of reinforcement is required to maintain an adequate proficiency level?

### 3. Monitoring Manual Control Actions

The pilot using a flight director or automatic system for control wants to spend a certain amount of time monitoring the confidence-inspiring situation information. This is how he gains and maintains confidence that all is going as expected. We speak of this time that he spends monitoring the situation information as his monitoring workload margin. It can be expressed either as a fraction of time, the dwell fraction, or as the fraction of the number of looks, the look fraction. Both the dwell fraction and the look fraction are obtained from eye point of regard (EPR) measurements, which are discussed subsequently in Subtopic E.

Sufficient monitoring margin is essential for the pilot to perceive exceedence of tolerances or specified values related to the task. Most of the pilot's status displays present the flight motion variables which are constituents of the automatic or flight director commands. Other status displays are common to engine or radar instrument monitoring, where the effects of manual control are not displayed. Still other status displays are common to traffic monitoring, where intervention for the purpose of control may be exceptional. This we shall call "monitoring and decision response" as discussed previously. More about measuring and interpreting this is presented in Ref. 1.

One purpose of the research reported in Refs. 51 and 52 has been to improve the bases for interpreting and predicting the partition of the pilot's time between the monitoring margin and the fraction of time

required for control. Estimates of average monitoring display threshold exceedence frequencies in terms of a level of pilot confidence in his situation, coupled with two conservation principles, viz., the conservation of look fraction and of dwell fraction, provide one basis for interpreting and predicting the partition of scanning workload for monitoring and control. The results of the partition provide estimates of the average scanning frequencies and dwell fractions for control as well as monitoring. The dwell fractions also represent the temporal probabilities of fixation. From these predictions, one can estimate the dwell intervals, look intervals, link values, and other scanning parameters desired (Ref. 53).

The detailed development of a simplified approximate method for partitioning the scanning workload required for monitoring and controlling a task with a single primary director display is given in Ref. 51 and with two primary director displays in Ref. 52 for a STOL approach. The properties of the pilot's scanning remnant and properties of the partition of scanning workload may conspire to compromise the pilot's confidence in his situation, to compromise his error performance, or both, so that his subjective impression of the overall task workload will be high.

The methods discussed so far rely on measurements of the pilot's scanning remnant in order to account for the potential role of parafoveal and peripheral vision in controlling and monitoring (e.g., Refs. 54 and 55). This is because one must be careful to distinguish between (measurable) eye movements and (unmeasurable) attention allocation between controlling and monitoring tasks.

A different approach to the real-time determination of human attention allocation between controlling and monitoring tasks is provided in Ref. 56. This approach uses an algorithm employing fading-memory system identification and linear discriminant analysis. The identification algorithm is used to determine the input-to-output relationship of the human operator in combination with the controlled element. A linear discriminant function is then used to detect identified parameter changes that indicate a shift in the operator's allocation of attention

(between controlling and monitoring) in excess of what is expected from a norm. The norm can be a running average of the discriminant as in Ref. 56 or could be based on a running average of the eye scanning measurements.

The authors conclude that the feasibility of the method in Ref. 56 depends on the control task being predominant and the monitoring task requiring infrequent attention. If events being monitored occurred frequently, the identifier did not adapt quickly enough and the relative measures of the discriminant function did not react appropriately. This may have been because the authors chose to subject the identified coefficients of the difference equation to discriminant analysis. Reference 57 shows that frequency domain measures are preferable to difference equation coefficients for representing identified parameter changes in a unique and sensitive way.

### 4. Monitoring Automatic Control

If we beg the question of the role for human intervention following detection of a failure during automatic landing, the results of measurements reported in Ref. 58 provide elapsed times for failure detection as functions of failure magnitude. The failures were restricted to glide slope and airspeed instrument failures, so that they did not affect the operation of the automatic landing system.

The fixed base simulation in Ref. 58 comprised the last five minutes of transport aircraft landing approaches starting on course at 2500 ft height and 10 miles from the runway threshold with fully automatic control. A high percentage (83 percent) of runs with single instrument failures was chosen to provide sufficient data for analysis of variance in a reasonable experimental interval of simulator occupancy. Obviously such a high failure rate is unrealistic and might bias the pilot to expect the failure. If full mission simulation and a more realistic failure rate had been employed, however, the effects of fatigue on the vigilance of the pilot might have confounded the results. (The authors include a compensating observation error threshold in the fitted model

of the pilot's decision function to correct for the a priori probability of failure in making realistic predictions.)

The participating pilots were told in advance that failures at random times would occur in either the airspeed or glide slope indicators, but that they should use other instruments for verification. There was no feedback to the pilot concerning his failure detection performance, however, because it was found in previous experiments (Ref. 59) that such feedback biased his next decision. His knowledge of a sequence of mistakes drove him to overcompensate with intense vigilance, and vice versa. When the pilot detected a failure, he pressed a button and the run was terminated. Otherwise, the run continued through touchdown, after which the pilot filled out a report in which he stated which instrument had failed and how he detected the failure.

The experimental results are interpreted with the aid of a fitted algorithmic model of the pilot as a monitor. The model includes a linear estimator and a decision rule. The linear estimator is a Kalman filter with measurement errors, rather than state estimates, as outputs. The decision rule is based on sequential analysis, but is modified for the special case of failure detection.

The use of the model for predicting absolute values of detection times depends on the limited experience in Refs. 58 and 59. In general, the pilots in both experiments preferred to operate at approximately equivalent but relatively low probabilities of false alarm and miss ($\cong 0.05$) with an observation error threshold between one-sixth and one-quarter of the observed standard deviation. These results need now to be compared with analogous results obtained under more realistic conditions to determine effects of crew fatigue on vigilance. Furthermore, experiments in monitoring automatic control must also consider the roles for human intervention after a failure has been detected as well as the effects of human participation in advance of a failure on vigilance. Before we discuss in the final topic some measurements which addressed this issue of the effects of participation on vigilance, we shall mention another theoretical treatment intended to help in interpreting measurements of the human operator's monitoring behavior.

Algorithmic techniques are also applied in Ref. 60 to develop two theoretical models for predicting human operator performance when monitoring an automatically controlled system. In one construct it is hypothesized that the operator monitors displays in order to detect failures most rapidly. In the other construct it is assumed that the displays are sampled in order to reconstruct the system status information in some sense which is optimal. In both cases the models employ a fractional value of attention to monitoring each displayed variable. These fractional values of attention are not necessarily measurable unless they can be correlated with eye scanning statistics to be discussed in Subsection E. Furthermore, the cost functionals employed in the respective optimization processes are not readily measurable either, unless subjective evaluations of the operator's strategy are used to assess the relative importance of costs.

The authors of Ref. 60 also discuss the relationship of their two theoretical models to existing prediction techniques for monitoring based on equal attention, peak excursion monitoring, and Nyquist frequency, for examples. The authors conclude that a weighted combination of failure detection and status estimation criteria offers the best potential for interpreting measurements of human operator monitoring behavior.

### 5. Monitoring Manual and Automatic Control

In our final topic of this section, we call the reader's attention to the measurements reported in Ref. 61, which examined the effects of the pilot's participation in the control task on his workload and failure detection performance during a simulated low visibility landing approach in a transport aircraft in turbulence. In these experiments the failures occurred in either the lateral or pitch axis of the flight control system so as to cause relatively slow drift in the course or flight path of the aircraft. Subtle failures, rather than hardover failures, were deliberately chosen to exercise the threshold of the pilot's failure detection capability. Sometimes the failure occurred in

an automatically controlled axis; other times, in a manually controlled axis.

The fixed base simulation in Ref. 61 began on the final approach course at a point seven miles beyond the outer marker and terminated either at touchdown or when a positive rate of climb had been established following the initiation of a go-around by the pilot. Failures were introduced randomly but only between the heights of 1800 and 800 feet (inside the outer marker). Although commercial transport landings were being addressed with airline pilots participating, the simulator did not incorporate all of the display and control capabilities necessary for Category 3 operations. Hence the authors elected to require a missed approach in the event that a pilot detected a failure. Thus the related issues of human intervention to correct, recover, and land were avoided, and failure detection time was adopted as the only measurement in the control failure experiments.

The experiments involved four levels of pilot participation in monitoring and controlling the aircraft:

a) Pilot monitoring all axes with autopilot controlling all axes

b) Pilot controlling only the lateral axis with autopilot controlling the pitch axis and autothrottle coupled

c) Pilot controlling the pitch axis and throttles with autopilot controlling only the lateral axis

d) Pilot controlling all axes.

Workload measurements were made in the absence of failures with the aid of a disjunctive reaction time measurement using a red warning light-cancelling subsidiary task. A workload index was computed in the manner of Ref. 62. Failure detection time measurements were made in the absence of the light-cancelling subsidiary task.

The workload measured with the pilot controlling the pitch axis and throttles (split axis participation Case c above) was over 50 percent greater than the workload measured with the pilot controlling only the

lateral axis (split axis participation Case b above). The workload index was approximately additive with respect to the manual control task.

The failure detection times in a manually controlled axis were significantly longer than detection times in an automatically controlled axis. Failures went undetected only in a manually controlled axis.

Detection times for lateral axis failures were significantly longer than for pitch axis failures at comparable levels of workload.

Higher levels of root-mean-square turbulence velocity resulted in higher levels of workload and longer failure detection times at comparable levels of pilot participation.

Since an increase in workload accompanied an increase in the level of pilot participation, the authors attempted to separate the effects of participation and workload on failure detection time. In fact, detection time did not increase monotonically with workload, thus suggesting that participation level did indeed influence detection time over and above the concomitant increase in workload. Nevertheless, increases in workload induced by turbulence without a change in level of pilot participation did increase detection time significantly.

Not investigated in this study and thus remaining a subject for research are the related issues of human intervention to recover and land, given that the necessary performnce monitors, fault annunciators, and flight control displays are provided. A variety of flight tests (e.g., Refs. 63, 64, and 65) have suggested that such successful intervention is possible.

## D. EVALUATING MEASURES WHICH REFLECT OPERATOR WORKLOAD

Workload motivates the human operator up to a point, where, in his judgment, either he experiences difficulty in maintaining the desired (or required) task performance by adapting his behavior, strategy, or technique or he believes he may no longer be capable of coping to a prescribed degree with an unexpected intrusion or failure. Operational

conditions such as these represent limits to the adaptability of an operator. In such limiting conditions, an operator is liable to err, and system performance is likely to degrade. Such operational conditions are said to impose high cognitive or perceptual-motor loading, which, for our purposes here, can be defined as the conscious involvement of the operator's corresponding systems in various tasks.

It has proven difficult to assess the compatibility of a man-machine system solely on the basis of a system performance decrement under cognitive and/or perceptual motor loading, because: (i) the human operator maintains a fairly wide workload margin, (ii) his homeostatic stability tends to attenuate measured variations in his performance and in his antonomic and somatic functions under stress, and (iii) variations in his cerebrospinal functions are even more difficult to measure and interpret. Furthermore, there are as yet no universal commensurate measures of the different types of loading among these functions which characterize the human operator nor among the different types of tasks which characterize national airspace operations. Consequently, other measures of operator loading have been used perforce.

The most common succcessful measure of workload has been subjective, viz., pilot opinion rating. Although of psychometric quality, these ratings are heavily weighted by an "expert's" introspective impression of the task loading and are more reliable as relative measurements when employed in comparative circumstances. Nevertheless the most common pilot opinion rating scales, the Cooper and Cooper-Harper scales, have acquired disciplined significance in rating flying qualities and are now commonly accepted as absolute measurements when rendered by trained experimental test pilots.

For discrete tasks in combination with more or less continuous control tasks, for supervisory control tasks with great latencies, and for most communication and navigation tasks, identifying and predicting detailed dynamic cognitive and sensorimotor behavior and associated workload are beyond current capabilities. In general, these types of tasks exhibit one or more of the following characteristics.

1. May need to be performed during high activity periods and can take a substantial amount of time.

2. May require extended cognitive activities without measurable response, including concentration, memory, logic, and/or referral to and correlation of supplementary data sources (for example, maps, charts, notes) for performance.

3. May precipitate a chain reaction of additional tasks into future time if not performed at the proper time on operator's initiative.

4. Can be performed incorrectly, omitted, or delayed for a significant time period after performance is required before it becomes obvious that something is wrong. Stated differently, it may require the operator to remember that at some specific future time he must perform some specific control functions.

Because the principles of these and other types of operator sensorimotor behavior and workload assessments are at the exploratory or low-confidence fringe of the theory of manual control, full mission simulation and empirical testing techniques must be employed. Among the objective measurements needed are those which are indicative of cognitive and perceptual-motor workload.

Various techniques have been developed for the estimation of the cognitive and perceptual-motor workload imposed upon the human operator of a complex vehicle (Refs. 66 through 74). We have partitioned these into the six basic categories and subsidiary techniques listed in Table 6.

Table 6 is arranged in approximate order of increasing complexity of measurement. A summary of the more relevant workload identification techniques is given by Ref. 74. A brief review and critique of each technique with references has also been given in Chapter VI of Ref. 75, and an updated annotated bibliography, using the topics of Table 6, is available as Ref. 69. We shall comment briefly on each major category in Table 6.

TABLE 6

TYPES OF COGNITIVE AND PERCEPTUAL-MOTOR
WORKLOAD MEASUREMENTS

A. Subjective Psychometric Ratings (supported by answers to questionnaires and by operator commentary)

B. Objective Workload Correlates

1. Auxiliary task techniques

.1 Auxiliary workload margin at a constant main task level of performance

.a Adaptive psychomotor task

.b Adaptive cognitive task

.2 Main task performance decrement at prescribed auxiliary task loads

.a Discrete-response auxiliary task

.b Forced scanning task

.c Multiaxis tracking and flying

2. Varying difficulty main task

.1 Sudden change in effective controlled element dynamics usually adverse

.2 Critical instability task

.3 Adaptive change in effective controlled element dynamics or difficulty

.4 Variable forcing function noise content at prescribed auxiliary task load

.5 Interrupted perception on main task and continuous tachistoscopy

3. Eye-point-of-regard measurements

.1 Scanning behavior patterns

.2 Scanning workload

TABLE 6 (Concluded)

4. Operator's dynamic behavior (e.g., describing function and remnant parameters)

C. Psychophysiological Correlates

1. Heart rate

2. Respiration rate

3. Neuromuscular tension

4. Evoked cortical potentials

5. Galvanic skin response

6. Pupillometric response

## 1. Subjective Psychometric Ratings

Subjective rating, such as given by the Cooper scale shown in Table 7 or by the modified Cooper-Harper scale in Fig. 13 (Ref. 76), is a direct workload index in that the actual mission can be performed without additional measuring equipment or tasks being required. The Cooper scale and Cooper-Harper scale (Ref. 77) are very nearly functionally psychometric (Ref. 78). The error introduced by averaging Cooper ratings, rather than their psychometric equivalent, is small provided enough trials have been made to ensure confidence in the ratings. The Cooper and Cooper-Harper scales are shown in Ref. 78 to be overly sensitive at the inferior ends, so that attaching significance to a difference of one Cooper unit between ratings at the inferior end would require a relatively large number of trials.

The state-of-the-art is well developed for making flying qualities ratings that are reliable and meaningful with respect to operational task demands and vehicle response characteristics. However considerably less work has been devoted to calibrating objective correlates of pilot workload in terms of pilot opinion ratings simply because few measurable workload indices have been available. Psychometric rating scales for task evaluation in terms of "controllability and precision" and "attentional workload" are presented in Table 8 from Ref. 55. Two scales for rating the usefulness of the status information and the amount of clutter in the display are also presented in Table 8 from Ref. 55.

## 2. Objective Workload Correlates

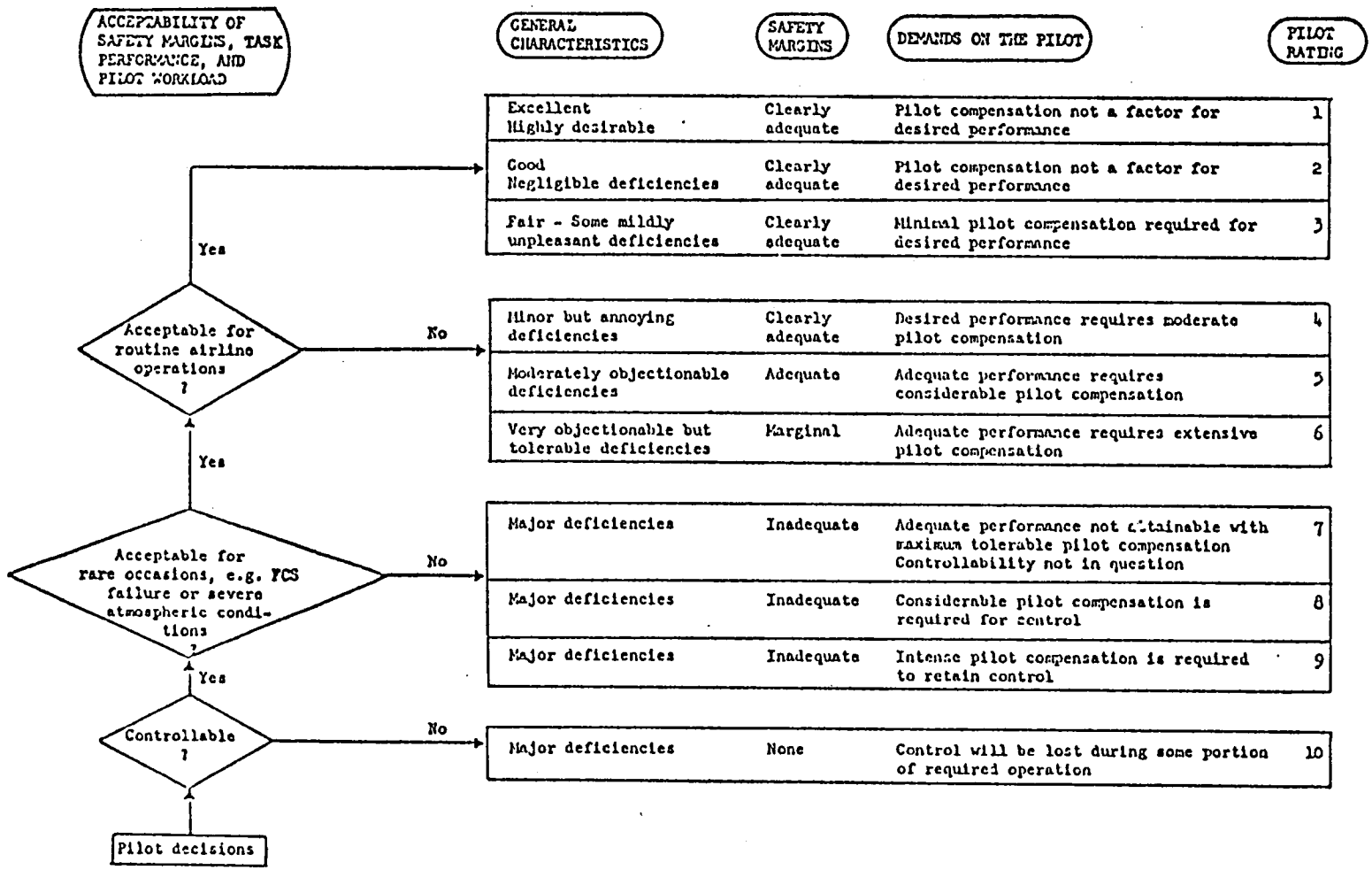
a. Auxiliary Task Techniques. By far the most common technique for controlling and measuring perceptual-motor loading is the use of auxiliary tasks of one type or another. The auxiliary task is intended to occupy the operator's reserve (or excess) capacity in one sensorimotor modality. However, it has been established that the reserve capacity measured in one sensory modality may not apply to other modalities. Therefore, it is vital that the sensorimotor modality of the loading

Table 7. The Original Cooper Scale (From Ref. 79)

| COOPER | | | | | PR |
|---|---|---|---|---|---|
| DESCRIPTION | ADJECTIVE RATING | MISSION | PRIMARY MISSION ACCOMPLISHED? | CAN BE LANDED? | |
| Excellent, includes optimum | Satisfactory | Normal operation | Yes | Yes | 1 |
| Good, pleasant to fly | | | Yes | Yes | 2 |
| Satisfactory, but with some mildly unpleasant characteristics | | | Yes | Yes | 3 |
| Acceptable, but with unpleasant characteristics | Unsatisfactory | Emergency operation | Yes | Yes | 4 |
| Unacceptable for normal operation | | | Doubtful | Yes | 5 |
| Acceptable for emergency operation (stab. aug. failure) only | | | Doubtful | Yes | 6 |
| Unacceptable even for emergency condition (stab. aug. failure) | Unacceptable | No operation | No | Doubtful | 7 |
| Unacceptable — dangerous | | | No | No | 8 |
| Unacceptable — uncontrollable | | | No | No | 9 |
| $9/*! Did not get back to report | Unprintable | What mission? | | | 10 |

# Figure 13

**Modified Cooper-Harper Rating Scale**




| GENERAL CHARACTERISTICS | SAFETY MARGINS | DEMANDS ON THE PILOT | PILOT RATING |
|---|---|---|---|
| Excellent<br>Highly desirable | Clearly adequate | Pilot compensation not a factor for desired performance | 1 |
| Good<br>Negligible deficiencies | Clearly adequate | Pilot compensation not a factor for desired performance | 2 |
| Fair - Some mildly unpleasant deficiencies | Clearly adequate | Minimal pilot compensation required for desired performance | 3 |
| Minor but annoying deficiencies | Clearly adequate | Desired performance requires moderate pilot compensation | 4 |
| Moderately objectionable deficiencies | Adequate | Adequate performance requires considerable pilot compensation | 5 |
| Very objectionable but tolerable deficiencies | Marginal | Adequate performance requires extensive pilot compensation | 6 |
| Major deficiencies | Inadequate | Adequate performance not attainable with maximum tolerable pilot compensation<br>Controllability not in question | 7 |
| Major deficiencies | Inadequate | Considerable pilot compensation is required for control | 8 |
| Major deficiencies | Inadequate | Intense pilot compensation is required to retain control | 9 |
| Major deficiencies | None | Control will be lost during some portion of required operation | 10 |

# TABLE 8 PILOT OPINION RATING SCALES

Rating Scale for Utility of Status Information

| Criteria | Descriptive Phrase | Rating |
|---|---|---|
| Usefulness* of the information supplied, on the specified display unit, on the vehicle status - expecially the relevant flight path vector states, such as: altitude, speed, heading attitudes, path errors, etc. | All desired states presented with adequate resolution and readability | S1 |
| | Many of desired states presented, with a few deficiencies in scaling, resolution or readability | S2 |
| | Some desired states presented, and/or some problems with scaling, resolution or readability | S3 |
| | Inadequate number of states, or serious deficiencies scaling, resolution or readability | S4 |
| *useful with respect to the mission phase, task criteria and operator's sense of vehicle safety. | No direct status information or unusable | S5 |

Rating Scale for Task Controllability and Precision

| Category | | Descriptive Phrase | Rating |
|---|---|---|---|
| Controllable? | Precise? | | |
| Yes | Yes | Very easy to control, with good precision | C1 |
| | | Easy to control, with fair precision | C2 |
| | No | Controllable, with inadequate precision | C3 |
| | | Marginally controllable | C4 |
| No | | Uncontrollable | C5 |

Rating Scale for Clutter

| Criteria | Descriptive Phrase | Rating |
|---|---|---|
| Degree of subjective symbol - background clutter on specified display unit | Completely uncluttered - e.g. only one pair of elements | K1 |
| | Mostly uncluttered - no confusing or distracting elements | K2 |
| | Some clutter - multiple elements competing for attention | K3 |
| | Quite cluttered - difficult to keep track of desired quantities among competitors | K4 |
| | Completely cluttered - nearly impossible to tell desired elements or quantities due to competing elements | K5 |

Rating Scale for Attentional Workload

| Criteria | Descriptive Phrase | Rating |
|---|---|---|
| Demands on the operator attention, skill, or effort | Completely undemanding and relaxed | D1 |
| | Mostly undemanding | D2 |
| | Mildly demanding | D3 |
| | Quite demanding | D4 |
| | Completely demanding | D5 |

task be representative of sensorimotor loading in the operational situation. Furthermore it is important to select an auxiliary task that has some relevance and face validity for the operator in the context of his customary and exceptional duties.

Kelley, Hudson, and others have developed the cross-adaptive input scheme (Refs. 80 and 81) for varying the difficulty of auxiliary tracking tasks to insure that a constant main task level of performance is maintained. In this type of scheme, the difficulty of the auxiliary task increases as long as the main task error is less than a criterion level, and vice versa for errors over the criterion. The asymptotic level of auxiliary task difficulty then provides a measure of the operator's excess control capacity with respect to the main task.

One of the most promising techniques for measuring excess control capacity is the cross-coupled adaptive subcritical tracking task described in Appendix B. In this technique the instability of the auxiliary task increases as long as the main task error is less than a criterion and vice versa for errors over the criterion. The asymptotic value of the instability is proportional to the operator's excess control capacity with respect to the main task*. As long as the operator's normal complement of tasks includes a tracking control task, it is usually possible to embed the cross-coupled adaptive subcritical tracking

---

* It turns out that the asymptotic value of the instability is an objective correlate of subjective rating and, in fact, from subjective ratings one can estimate the excess control capacity via the calibration of the objective correlate in terms of subjective rating (Ref. 12). In many cases the measurement of excess capacity need not be made! Nevertheless we need more extensive calibrations of the objective correlate in terms of subjective rating, including some which demand a level of skill development higher than compensatory and which involve more than a single operator.

Workload is monotonically related to excess control capacity, attentional demands, and ability to cope with the unexpected. All three of these can be measured objectively for situations where a subjective assessment of cognitive (e.g., search and recognition, monitoring, decision making, etc.) and/or control tasks can be found. These calibrations between subjective and objective measures are thereafter used to quantify the workloads without having to resort to elaborate, time-consuming, and sometimes non-realistic objective procedures.

task among the operator's tasks with high face validity (for example, see Ref. 82). If on the other hand, the operator does not customarily perform a suitable tracking control task, or there appears to be no valid way to embed the cross-coupled subcritical tracking task with high face validity, it may be possible to embed an auxiliary cognitive task instead.

An auxiliary cognitive task involving item recognition, also described in Appendix B, can often be embedded among the visual or auditory commands and voice "traffic" reaching an operator. Moreover an item recognition task can even be adaptively cross coupled to the operator's error performance on his primary task to avoid encroachment so that the resulting measurement will more accurately reflect reserve capacity (Ref. 83). Details of this adaptive cross-coupling are also described in Appendix B.

Using one or more of the auxiliary task techniques described above could be extremely valuable for increasing the effectiveness of the proposed full-mission simulation. One possible scenario would be to induce fatigue by having a flight crew fly a part-task simulator that was configured with an auxiliary task(s) prior to flying the full-mission simulator. The same part-task simulator could be used to simulate aircraft interacting with that employed in the full-mission simulation and flown by alternate crews.

b. Varying Difficulty Main Task. The leading contender for the continuous type of main task loading is the use of a critical instability task, as described in Refs. 84-87.

For operational situations involving failure management, an ordinary continuous auxiliary task loading is not appropriate. A progressively degraded main task or possibly an unexpected change in controlled element properties would be better (Ref. 48).

Reference 88 has also successfully employed a variable disturbance forcing function for the main task by regulating a prescribed auxiliary task load. Reference 89 has employed interrupted perception on the main task to vary its difficulty.

c. Eye-Point-of-Regard Measurements. This is the original measure of pilot fatigue proposed by McGehee (Ref. 90) and evolved by Fitts, et al (Refs. 91-102). If gross inconsistencies with the display arrangement hypothesis (Ref. 103) are observed on a display arrangement, scanning and eye traffic measures are indicators of abnormal distributions of scanning workload. However, these are not absolute measures and are useful only in comparing the partition of scanning workload among different display arrangements. In connection with integrated displays, measures of scanning and eye traffic may suggest phenomena like "stare mode" or "tunnel vision." In a stare mode, fixing the eye-point-of-regard serves to stabilize the eye for good parafoveal viewing, and the measured fixation point may be unrelated to the information actually being used. Conversely, "tunnel vision" without scanning may exclude perception of some parafoveal signals needed for multiloop control. Additional measures such as the describing function might be required to resolve the ambiguity between these two phenomena. Eye-point-of-regard measurments will be discussed more fully in Subsection E.

d. Measured Pilot Response Characteristics. The value of measured pilot response properties such as the adaptive parameters (gains, lead, lag, effective time delay) fitted to the pilot's describing function, system stability margins, and pilot remnant properties, lies in their empirical correlations with high workload situations. For example, we know that the requirements for generation of lead-time constants in excess of 1 sec are considered high workload tasks by pilots. The increment in effective time delay that accompanies low frequency lead generation has in the past been considered a cause of perceptual-motor load. Reciprocal effective time delay as a function of the order of low frequency lead equalization is shown in Fig. 14. These parameters have been correlated primarily with handling qualities ratings and not with perceptual-motor load measures as such. The component of effective time delay which is related principally to neuromuscular tension provides one of the clearest examples of an association between a measure of pilot response which is known to demand higher subjective workload and a physiological measure. Figure 15 shows that the average effective time

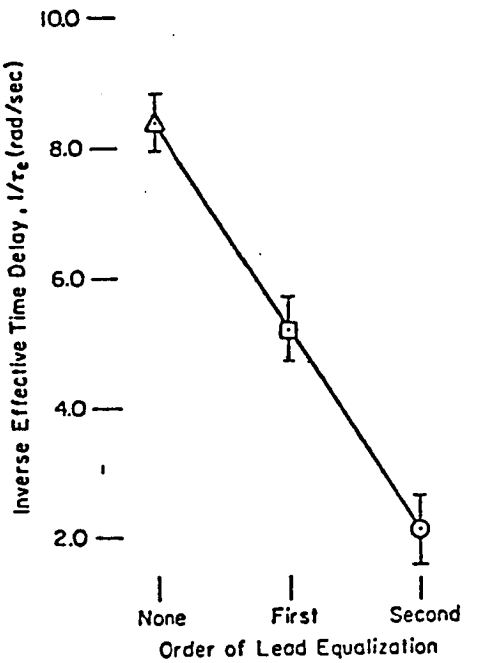



Figure 14. Inverse Effective Time Delay as a Function of the Order of Lead Equalivation Required of the Pilot (from Ref. 104)

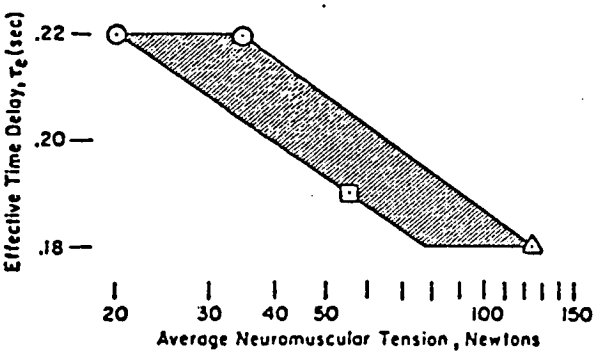



Figure 15. Effective Time Delay as a Function of Average Neuromuscular Tension

delay decreases as average neuromuscular tension increases. Such correlations need to be established before one can predict pilot response properties to meet the task demands using multiloop feedback theory and established pilot adaptation rules.

### 3. Psychophysiological Correlates

Many workload measurement schemes include additional measurements on the pilot-vehicle system. A battery of psychophysiological measurements is very attractive because such measurements can easily be made during the performance of the actual or simulated task and do not require auxiliary tasks to provide a score. The basic assumption is that the physiological variables are in some way correlated with the workload of the task at hand. These correlations have not yet been firmly established and the interrelationships among them are only beginning to be understood (Ref. 62, 105, and 106). The most popular measurements are those related to the cardiovascular and respiratory systems: heart rate and its variation on a beat-to-beat basis (sometimes called heart acceleration); various measures of pulse pressure, breathing rate, depth of breathing, tidal volume, and so on. Measurements of neuromuscular involvement include filtered absolute electromyogram levels in both the active and passive limbs, integrated absolute electromyograph from a series of sites, grip pressure, neuromuscular tremor frequencies, etc. In certain cases there are strong correlations between physiological measurements. For example, in the resting state there is a periodic psychophysiological fluctuation in the heart rate called "sinus arrythmia," which often correlates with the periodicity of breathing. Under high perceptual-motor loading conditions the sinus arrythmia tends to vanish, while the average heart rate tends to elevate somewhat. Preliminary data from continuous tracking with a subcritical unstable controlled element suggest that the change in sinus arrythmia amplitude accompanies higher bodily neuromuscular tension levels. Kalsbeek has also found an analogous attenuation in sinus arrythmia under ADT stress (Ref. 107).

Some measures of more emotional involvement include a number of variations of galvanic skin response (GSR best exemplified by palmar skin resistance), eye pupil diameter, and local temperature fluctuations at selected skin sites. There is evidence that pupillometric fluctuations and sudden decreases in palmar skin resistance accompany systemic

pulsations in neuromuscular tension that seem to follow "arming" changes in perceived signals.

One of the few measurements presumably directly related to mental activity is electroencephalogram (EEG). However, just what combination of sites and what signals best indicate perceptual-motor loading has not been determined. The most common indicator of awareness is taken to be the changes in the alpha-rhythm component of the EEG signals that at least show an observable correlation with certain visual and mental activities. Such measurements are very popular in the USSR (Refs. 108 and 109), in the Netherlands (Refs. 107, 110, and 111), and in England (Ref. 112). In the United States, Roman has collected in-flight measurements during simulated and real missions (Ref. 113).

We have examined and selected a number of psychophysiological measurements for investigation in NASA-sponsored critical task research. Based upon a survey of the literature and consultation with a number of researchers in the field, those measurements that appear to be most relevent are: instanteneous heart rate and acceleration, respiration rate and acceleration, depth of breathing, palmar skin resistance, passive limb EMG, grip pressure, and eye blink rate. Fairly standard techniques are available for all of these measurements, and they lend themselves to either simulator or in-flight situations.

## E. EYE POINT OF REGARD (EPR) MEASUREMENTS

As mentioned in the previous section, EPR measurements can be used to obtain the pilot's monitoring workload margin while performing either manual or automatic tasks. EPR measurements are also used for other purposes in conjunction with flight control and monitoring tasks, some of which are discussed in this subsection. The two subsequent topics discuss problems in reducing raw EPR data and future applications, respectively.

## 1. Background

A summary of eye movement studies in flight control and monitoring tasks is contained in Ref. 108 from which the following is extracted directly:

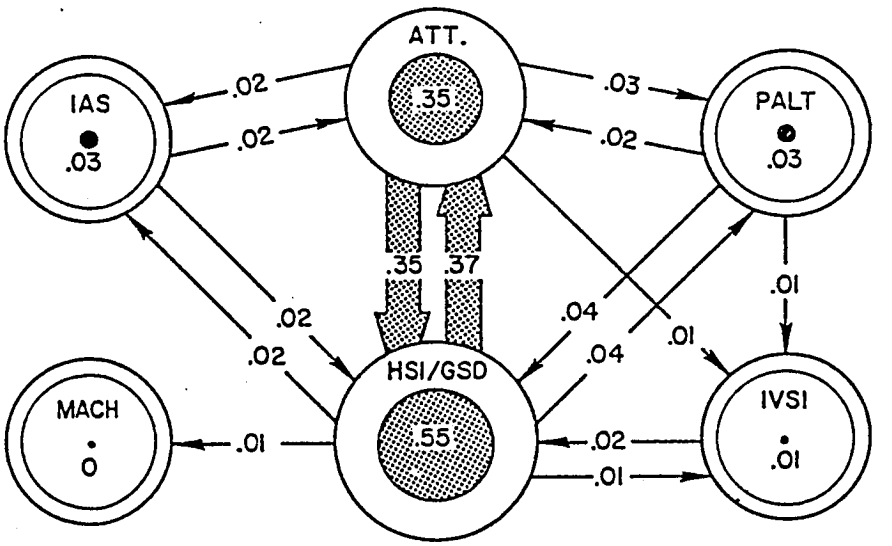
> "The inspiration for much of this [prior] eye movement work was founded on the belief that the cues used by the pilot in controlling flight would be revealed by noting the (separated) instruments upon which the fovea of the eye was fixating inside the cockpit under instrument flight rules, and by correlating the directions of fixations external to the cockpit with signficant ground-based cues in landing approaches under visual flight rules. Information about the useful instrument flight control cues was believed to be fundamental to an understanding of the function served by flight instruments. It was expected that this understanding would, in turn, form a basis for improving the design of aircraft instruments, increasing the efficiency of instrument flight training, and simplifying the task of instrument flying.
>
> "Today we are still working to fulfill this expectation, because the premise on which it was founded twenty years ago has been shown to be only a partial truth for several reasons. Pilots develop an ability to operate effectively on parafoveally and peripherally perceived information (Ref. 115), albeit with some limitations (Ref. 116), and, of course, on reinforcing (i.e., nonconflicting) motion and aural cues. Further, there is considerable indirect evidence (e.g., Ref. 117) that in 'stare mode' circumstances fixing the eye-point-of-regard serves merely to stabilize the eyeball for good parafoveal viewing, so that the fixation point may be unconnected with the information actually used, or even perceived, by the pilot. We cannot say that what is being fixated necessarily corresponds to an input.
>
> "The inspiration for the earliest pilots' eye movement studies — that scan patterns might be useful for workload measures — was revived more recently in Ref. 118. While scan patterns are indeed relevant to workload, the connection is not simple. The eye requires fixation to keep the eyeball stable, so there is a kind of Parkinson's

law for the eyeball — the sum of the fixation dwell times on the instruments expands or contracts to equal the time available (neglecting saccadic times). There is, of course, a minimum dwell time of about 0.4 sec per instrument, so it is possible to contrive saturated conditions where the control task demands pilot fixations on too many instruments too often in order to maintain control. But the interpretation of such results would often be ambiguous if one is looking for the pilot's inputs.

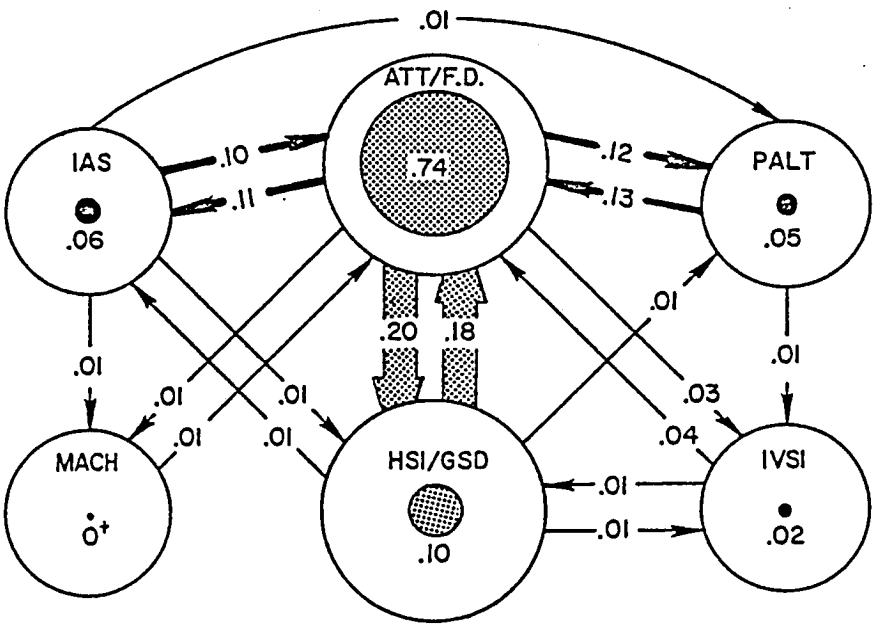
"The principal cost of the pilot's scanning behavior is an increased 'remnant.' This depends on the sampling frequency, fixation dwell time, and sampling frequency variations, as well as the observed signal variance. The remnant represents pilot control movements which are incoherent, i.e., not linearly correlated (via the describing function) with the externally imposed forcing functions. The remnant acts like an injected noise, and is the real cause of saturation in multi-instrument displays. So, as we said at the outset, measurement of eye fixation is certainly connected with pilot inputs and workload but the connection is by no means simple."

A sample of the type of data that can be inferred from EPR measurements is shown in Fig. 16 (adapted from Ref. 119). The instruments shown in this figure and their positions relative to one another are representative of most conventional jet transports. The numbers within instruments shown in Fig. 16 are called the "dwell fractions," which represent the proportion of the total time during which fixations dwell on a particular instrument. Since the cumulative sum of all dwell fractions, including blinks and distractions, must equal unity, by definition, the dwell fraction is also termed "fractional scanning workload" or "probability of fixation."

The numbers between the arrows shown in Fig. 16 are called the "one-way link-values," which are the proportion of all fixation transitions which go in the specified direction between a pair of instruments. The sum of the two one-way link-values between a pair of instruments is called the "two-way" link value. In 1950, new research extended the display arrangement hypothesis of 1944 to suggest that the pattern of link-values between instruments is indicative of the goodness of




*Manual ILS CONFIGURATION, CI*




*Flight Director Configuration, EI*

Figure 16. Measured Dwell Fractions and Transition Link Fractions for Manual and Automatic Apprcaches (from Ref. 119)

different panel arrangements. Since, in point of fact, the scanning statistics are quite stationary over measurement intervals as short as 100 sec, different one-way link-values between the same pair of instruments are also indicative of determinism in scan patterns. If the pilot's scanning behavior were represented by a truly random process (i.e., there was no deterministic "pattern") then the one-way link-values would be of equal magnitude. The results in Ref. 119 show no evidence of circulatory determinism in the scanning statistics. This simplification proves useful in making predictions of scanning behavior (Ref. 104).

## 2. Reduction of EPR Data

Widespread use of eye-point-of-regard data has always been hampered by the large amount of time required to reduce and process the raw EPR data. Because of this only a small fraction of the large amounts of EPR data recorded are ever used. Some of the general problems encountered in reducing raw EPR data, independent of the method used to record it, are discussed below.

The raw data for modern EPR measurement systems (e.g., Ref. 120) are usually available in the form of voltages that are proportional to the displacement of the fixation point in the visual field. In the past, these voltages have been recorded on strip charts and then manually reduced at the end of the experiment (e.g., subject looking at Instrument 5 for 2.3 sec, etc.). It turns out that the human analyst is extremely efficient at filtering out artifacts present in the raw EPR data but the turn-around time is long. Also, boredom probably causes a certain amount of error in the data reduction.

The raw EPR voltages could be converted to digital signals and sent to a computer which could, theoretically, be programed to process the raw EPR data automatically. Getting the raw EPR data into a computer is not a problem, but designing an algorithm that will properly reduce the EPR data has, to date, frustrated some researchers (e.g., Ref. 121). Some of the artifacts in the raw EPR data that cause problems are discussed below.

a. Noise. There are two sources of noise in the raw EPR data. First, the eyeball is constantly moving in order to create a stable image. Thus, even though a subject may be fixating on a single point in the visual field, the EPR measuring system will detect "movement." Second, the EPR measuring system itself may cause noise due to the method used to obtain the EPR voltages. The data reduction algorithm must reject both sources of noise.

b. Blinks and "Glitches." When a subject blinks it usually produces a definite and fairly repeatable pattern in the EPR signals (e.g., for the STI EPR system blinks appear as a quick look down and to the left). "Glitches" look like drop-outs in the data and are probably due to artifacts in the measuring equipment. The patterns produced by both glitches and blinks are easily recognized by the human analyst, but it is difficult to program a computer to recognize and correct these patterns.

c. Saccades and Fake Looks. A saccade, a quick jump in the point of regard, occurs when the EPR is in the process of transitioning from one instrument to another. The EPR signal, however, will appear to slew across the visual field, rather than immediately jump from one point to the next. Also, a "fake look" to a point in the visual field can result when the subject is transitioning from Point A to Point B and passes, but does not dwell, over Point C. As with blinks and glitches the saccades and fake looks are fairly easy patterns for the human analyst to recognize but it can be difficult to devise a computer algorithm to recognize them.

Other artifacts in the data due to the particular EPR system being used may also be present and must be considered if the algorithm is to be successful in automatically reducing the EPR data. For example, the STI EPR system uses the eyelid to detect indirectly the vertical movement of the eye. This unfortunately contaminates the EPR data with the eyelid dynamics, which appear to be nonlinear.

Even though the problem is difficult, as elucidated above, it is believed that a successful algorithm to reduce EPR data automaically can be developed. An algorithm for the STI EPR system has been developed but to date has not been programmed and tested with actual EPR data.

## 3. Future Applications

Future applications of EPR data will be dependent, at least to some degree, on the success of developing algorithms to reduce and process the raw EPR data automatically. The following partial list of future applications assumes that this capability is readily available.

a. Error Detection. How long does it take to detect an error condition? Is the error condition confirmed by cross checking? If so, what information is used to confirm the error condition? How long is it from the time when the error is detected until the time when corrective action is taken?

b. Emergency Action. What information is being used, or perhaps misused, in an emergency?

c. IFR to VFR Transition. How much cross checking of head-down instruments is done after "runway-in-sight?" What head-down instruments are used?

d. Display Optimization. Although this is not a "new" application it will continue to be a future application of EPR data, especially as it becomes easier and cheaper to process the raw EPR data.

e. Decision-Making Identification. What information is being used to make comlex decisions? Can EPR data be helpful in combined decision making and control strategy identification techniques?

f. Control Behavior Identification. A tacit assumption of current methods used for identifying pilot control strategy is knowledge of what the pilot is looking at. This is especially true of multiloop control tasks where the pilot control strategy is not always unique. Direct correlation of EPR data and control activity would be useful in these more complex control tasks. EPR data has already been correlated with measurements of the pilot's remnant in several experiments (Refs. 54 and 55) with favorable results which demonstrate the reality of scanning remnant as a cause of saturation in multi-instrument flight tasks. EPR data may also provide insight into latent control activity and the phenomenon of control reversals in flight simulators.

## SECTION VI

## CONCLUSIONS

A wide variety of proven measurements and data-reduction techniques suitable for identifying human error are recommended for use in connection with the NASA-Ames Research Center Man-Vehicle Systems Research Simulator Facility. Most of the measurement techniques are sufficiently unobtrusive that they do not interfere with either full mission simulation experiments or the operation of the simulator facilities. Many of the measurements will provide reduced data in situ for timely evaluation while an experiment is in progress. These and other measurements are also appropriate for describing ensembles of data in those instances where probabilistic generalizations may be justified after the experiment has been concluded.

Examination of the definitions, types, and sources of human error from Ref. 1 which need to be identified suggests that the classes of measurements indicated in Table 9 and further elaborated in Table 10 will distinguish certain types among the corresponding groups of human errors listed. Notice, however, in Table 9 that a particular class of measurements is capable of identifying more than one type of error. For this reason interpretation of a variety of measurements may be required to identify a particular source or type of error. In this respect the additional clues provided in Tables 9 through 11 in Appendix A hereto may be especially useful in helping to interpret system performance-centered and operator-centered measurements. Tables 12 and 13 in Appendix A, are designed to assist in the more difficult problem of identifying causes of error leading to inappropriate organization of perception and behavior at the executive level of the operator's activity-supervising control. This level of activity transcends the operator's various directly involved systems, such as the perceptual, cerebrospinal, autonomic and neuromuscular systems about which particular measurements can be made.

TABLE 9

MEASUREMENTS FOR IDENTIFYING HUMAN ERROR

| Basic Sources of Error (from Ref. 1) | Groups of Distinguishing Measurements: System Performance-Centered Measurements: Location in state space and time with respect to authorized boundaries and schedules | Location in state space and time with respect to critical limits | Command-following or disturbance regulation bandwidth or latency | Stability | Operator-Centered Measurements: Describing functions | Remnant | Opinion ratings | Workload | Eye point of regard | Psychophysiological correlates | Location in state space and time with respect to acceptable limits | Procedure-Centered Measurements: Time line analysis of discrete stimuli, responses, sequences, and latencies in procedures |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Extreme command or disturbance amplitude | X | X | X | X | X | X | X | X | | X | X | |
| Extreme command or disturbance bandwidth | X | X | X | X | X | X | X | X | | X | X | |
| Controlled-element change | X | X | X | X | X | X | X | X | | X | X | |
| Reduced attentional field in single channel operations | | | X | | X | | X | | | | | X |
| Diverted or divided attention and perceptual scanning in multi-input operations | | | X | | X | X | X | X | X | X | | X |
| Reversals | | | | | | X | | | | | X | X |
| Illusions, kinetosis | | | X | | | X | X | | | X | X | |
| Spontaneous improper actions | | | | | X | X | | | | | X | X |

TABLE 10

QUANTITATIVE AND QUALITATIVE MEASUREMENTS AND EVALUATION CRITERIA

PROCEDURE-CENTERED (Comparative evaluation criteria are based on standard pre-experimental time line analyses for the scenario)

Evaluation of discrete stimuli, responses, sequences, and latencies in time domain among the normal and emergency procedures involved in the following activities:

- Supervising and executing checklists
- ATC clearance compliance and reporting
- Execution of the flight plans and alternates, including flight profile management and use of change-over points
- Communication
- Navigation
- Identification
- Book-keeping, record-keeping, document and library management
- Aircraft systems operation (e.g., propulsion, fuel, electrical, hydraulic, wheels, brakes, auxiliary power, anti-icing, and environmental radar)
- Flying, i.e., guidance and control; manually and automatically
- Tactical decisions
- Overall crew supervision, management, and integration

SYSTEM PERFORMANCE (Evaluation criteria are commensurate with metrics and absolute in value)

Stability (e.g., phase or gain margins)
Command-following frequency bandwidth or temporal latency
Disturbance regulation bandwidth or latency

Location along flight plans/profile:

- Location in state space and time with respect to authorized boundaries and schedules, including unauthorized ground proximity

Propulsion:

- Location in state space with respect to critical limits

Structural load factors with respect to critical limits
Aerodynamic stall margins
Weight and center of gravity with respect to critical limits

TABLE 10 (Continued)

SYSTEM PERFORMANCE (Evaluation criteria are commensurate with metrics and absolute in value) (cont.)

At Approach Window:

- Location in state space with respect to window boundaries
- Probability of Approach Success

At Touchdown:

- Longitudinal and Lateral Touchdown Location with respect to runway
- Sink Rate
- Sideslip
- Heading
- Pitch, and Roll Attitudes
- Airspeed Error

Composite Measures

SAFETY MEASURES

Probabilities (Evaluation criteria are commensurate)

- Successful Landing
- Successful Missed Approach
- Accident or Incident
- Margin (Stall, performance, etc.)

Qualitative Assessments (Evaluation criteria are relative and subjective; the graceful degradation hypothesis provides a guide)

- Missed Approach Procedures
- Failure Detection Procedures
- Emergency Takeover Procedures

OPERATOR-CENTERED PERFORMANCE AND ACCEPTANCE MEASURES

Operator Dynamic Behavior (Evaluation criteria are relative)

- Describing Functions and Remnant (loops closed and equalization demanded; control-display associations and residual cross-coupling; sensitivity of stability, disturbance regulation, and command-following performance to variations in gain, time delay, and equalization; the adaptive feedback selection hypothesis and successive organization of perception hypothesis provide guides)

TABLE 10 (Concluded)

OPERATOR-CENTERED PERFORMANCE AND ACCEPTANCE MEASURES (cont.)

Operator Dynamic Behavior (evaluation criteria are relative) (cont.)

- Eye-Scanning Activity Distributions (incoherence in system performance caused by scanning remnant; system status monitoring threshold for confidence and decision making; the display arrangement hypothesis provides a guide.)
- Opinion Ratings (psychometric scales)
- Workload and Operability Assessment (excess control capacity; auxiliary task scores and loads; psychophysiological correlates; there is no guide to evaluation other than sensitivty and relative differences)

Psychophysiological Correlates (Evaluation criteria are subjective and relative)

- Heart rate and acceleration
- Respiration rate and acceleration
- Depth of breathing
- Palmar skin resistance
- Passive limb electro myography
- Grip pressure
- Eye blink rate

Operator Acceptance of System Performance

- Attitude, Attitude Rate, and Load Factor Variances from Trimmed Values (Evaluation criteria are commensurate and absolute, e.g., probabilities of exceeding acceptable levels from trimmed values)
- Control Displacement and Rate Variances from Trimmed Values (Evaluation criteria are commensurate and absolute, e.g., probabilities of exceeding maximum authorities)
- Response Compatibility and Motion Harmony--Automatic and Flight Director versus Manual Control (Evaluation criteria are relative to the response and motion attributes under manual control)
- Command Consistency--Flight Director versus Manual Control (Evaluation criteria are based on the consonance between the spectral distribution of status variables in the director command and the displayed stqtus variables themselves)
- Qualitative Assessments
  - Operator Commentary (evaluation criteria are subjective and relative).

Table 10 elaborates first on the measurements for procedure-centered evaluation. These are primarily discrete stimuli, responses, sequences thereof and latencies therefor in the time domain among the normal and emergency procedures involved in the listed activities. The comparative discrete evaluation criteria needed to identify human errors must be based on thorough pre-experimental time line analyses for the scenario. It is preferable to institute automatic recording of discrete activities by the crew members wherever possible. Thereafter to detect errors it is possible to employ automatic comparison of the recorded time-line of discrete activities with the pre-experimentally recorded time line of "normal" and "emergency" procedures established for the scenario.

Quantitative and qualitative system-centered performance measures and evaluation criteria are also listed in Table 10. Foremost among these are stability, command-following bandwidth and distance regulation bandwidth. Other system performance measures are ordinarily in the form of exceedences, means and variances since the major inputs of concern are random or can be considered such. The composite measures might be appropriate combinations of touchdown or window variables for example. The primary quantitative safety measures are expressed in probabilistic terms for commensurate evaluation. These are determined using the system performance measures (or, more precisely, their distributions) and the limiting factors of the scenario. Again for the approach and landing situation, examples might be Category II "window" sizes and landing gear limits. The assessments associated with safety are determined by evolving scenarios for missed approach, failure detection and emergency takeover procedures, wherein the crew's ability to control the failed system is considered.

Table 10 concludes with an elaboration of operator-centered performance and acceptance measures which serve as diagnostic aids for detecting human error. Foremost among these, because of their proven reliability, are operator-describing functions and remnant, eye-scanning activity distributions, and subjective opinion ratings. The opinion rating provides an overall operator-centered assessment of the total system. It is based on the qualitative assessment of workload and the

operator equalization demanded by the multioperator management and control structure. It should be supported by objective workload and operability assessment (for which opinion rating is the best calibrator), by the various listed measures which confirm operator acceptance, and finally by operator commentary.

For the procedure-centered human error data in Table 10 and other low probability events such as accidents or incidents, listed under "Safety Measures" in Table 10, we can usually depend on full mission simulation only for anecdotal and qualitative evaluation as in Ref. 5. Any statistical measures of confidence in procedural errors and other low probability outcomes would require months of accumulated experience at enormous cost. The outlook is much more favorable, however, for acquiring statistical measures of confidence in certain system-centered and operator-centered parameters from short-term temporal ensembles, where the ergodic hypothesis is reasonably valid. In this regard, system command-following bandwidth or latency, disturbance regulation bandwidth or latency, stability margin, and operator describing functions qualify from Table 10.

Part-mission simulation offers economy in the investigation of human error by virtue of its ability to focus on a particular flight segment (e.g., approach and landing) without spending resources on portions of the flight (e.g., cruise) of lesser interest or in which fewer errors might be expected. Repeated simulation runs by one crew or an ensemble of simulations involving many crews become quite feasible.

The possibilities for improper execution of the myriad of normal and emergency procedures within a particular flight segment can be examined in more detail in advance for part-mission simulation, simply because the volume of alternative possibilities is reduced by comparison with that volume in full mission simulation. Thus one is more likely to be prepared in advance with the necessary alternative detailed procedural time line analyses for comparing and judging the discrete stimulus-response activities to detect procedural errors in part-mission simulation.

Planning data collection beforehand specifically for the anticipated data reduction and statistical analyses is a general requirement for studies of human behavior. A significant investment of time and effort beforehand will assure more productive results from the measurements obtained in the actual experiment. In addition to ensuring that the assumptions required for the analyses are met, consideration of the fiducial statistical tests provides guidance in deciding how much data to collect. In some cases, evaluation of the power of a proposed test for detecting expected differences may lead to abandoning a measurement or even abandoning the experiment!

## REFERENCES


1. McRuer, Duane T., Warren F. Clement, and R. Wade Allen, A Theory of Human Error, Systems Technology, Inc., Technical Report No. 1156-1, May 1980.

2. Clement, Warren F., Human Error Classification and Recommendation of Critical Areas that Require Research, Systems Technology, Inc., Working Paper No. 1156-2, July 1980.

3. Billings, Charles E., John K. Lauber, Hallie Funkhouser, E. Gene Lyman, and Edward M. Huff, NASA Aviation Safety Reporting System Quarterly Report Number 76-1, April 16, 1976 Through July 14, 1976, NASA TM X-3445, 1976 (et seq.).

4. Anon., Man-Vehicle Systems Research Facility Experiment Planning Workshop, Summary Report, Ames Research Center, Moffett Field, California, MVSRF 105, 4-6 December 1979.

5. Ruffell Smith, H. P., A Simulator Study of the Interaction of Pilot Workload with Errors, Vigilance, and Decisions, NASA TM 78482, January 1979.

6. Imrich, Thomas, Concept Development and Evaluation of Airborne Traffic Displays, Flight Transportation Laboratory Report R71-2, Massachusetts Institute of Technology, June 1971.

7. Sanders, A. F., "Some Aspects of the Selective Process in the Functional Visual Field," Ergonomics, Vol. 13, No. 1, 1970, pp. 101-117.

8. Johnson, W. A., and Hoh, R. H., Determination of ILS Category II Decision Height Window Requirements, NASA CR-2024, May 1972.

9. Hofmann, L. G., et al., Investigation of Measuring System Requirements for Low Visibility Landing, AFFDL-TR-71-151, December 1971.

10. Anon., Standard Performance Criteria for Autopilot/Coupler Equipment, Radio Tech Commission for Aeronautics, RTCA, SC-79, 14 March 1963.

11. Personal communication between L. G. Hofmann and R. Monteleone, Sikorsky, 29 July 1970.

12. McRuer, Duane T., and Ezra S. Krendel, Mathematical Models of Human Pilot Behavior, AGARDograph No. 188, January 1974.

13. Jewell, Wayne F., and Ted M. Schulman, A Pilot Control Strategy Identification Technique for Use in Multiloop Control Tasks, NASA CR-152374, August 1980.

14. Elkind, Jerome I., Further Studies of Multiple Regression Analysis of Human Pilot Dynamic Response: A Comparison of Analysis Techniques and Evaluation of Time-Varying Measurements, ASD-TDR-63-618, March 1964.

15. Hoffman, Lee Gregor, Paul M. Lion, and John J. Best, Theoretical and Experimental Research on Parameter Tracking Systems, NASA CR-452, October 1965.

16. Astrom, K. J., and P. Eykhoff, "System Identification — A Survey," Automatica, Vol. 7, 1971, pp. 123-162.

17. Isermann, R., Prozessidentifikation-Identifikation und Parameterschätzung dynamischer Prozesse mit diskreten Signalen (Process Identification - Identification and Parameter Estimation of Dynamic Processes with Discrete Signals), Springer-Verlag, Berlin, 1974.

18. Gustavsson, I., L. Ljung, and T. Soderstrom, "Survey Paper," Automatica, Vol. 13, 1977, pp. 59-75.

19. McRuer, Duane, "Human Dynamics in Man-Machine Systems," Niemi (Ed.), A Link Between Science and Application of Automatic Control, Proceedings of Triennial World Congress of IFAC, Vol. 4, Pergamon Press, Elmsford, New York, 1979.

20. Kleinman, D. L., and S. Baron, Manned Vehicle Systems Analysis by Means of Modern Control Theory, NASA CR-1735, 1971.

21. Kleinman, D. L., and S. Baron, "A Control Theoretic Model for Piloted Approach to Landing," Automatica, 9, 339, 1973.

22. Kleinman, D. L., S. Baron, and W. H. Levison, "An Optimal Control Model of Human Response, Parts 1 and 2," Automatica, 6, 1970.

23. McRuer, Duane, Dunstan Graham, Exra Krendel, and William Reisener, Jr., Human Pilot Dynamics in Compensatory Systems Theory, Models, and Experiments with Controlled Element and Forcing Function Variations, AFFDL-TR-65-15, July 1965.

24. Hoh, Roger H., Samuel J. Craig, and Irving L. Ashkenas, Identification of Minimum Accpetable Characteristics for Manual STOL Flight Path Control, Volume III: Detailed Analyses and Tested Vehicle Characteristics, FAA-RD-75-123, III, June 1976.

25. Stapleford, R. L., S. J. Craig, and J. A. Tennant, Measurement of Pilot Describing Functions in Single-Controller Multiloop Tasks, NASA CR-1238, January 1969.

26. Weir, D. H., and D. T. McRuer, Pilot Dynamics for Instrument Approach Tasks: Full Panel Multiloop and Flight Director Operations, NASA CR-2019, May 1972.

27. Taylor, Lawrence W., Jr., "A Look at Pilot Modeling Techniques at Low Frequencies," Proceedings of the Sixth Annual Conference on Manual Control, Air Force Institute of Technology/Air Force Flight Dynamics Laboratory, 7-9 April 1970.

28. Rice, S. O., "Mathematical Analysis of Random Noise," Bell Systems Technical Journal, Vol. 23, July 1944, pp. 282-332; Vol. 24, January 1956, pp. 46-156. (These articles also appear in Selected Papers on Noise and Stochastic Processes, N. Wax (Ed.), Dover Publications, New York, 1954, pp. 133-294).

29. Blackman, R. B., and J. W. Tukey, The Measurement of Power Spectra, Dover Publications, Inc., New York, 1958, p. 19.

30. Gagne, Gilbert A., and Walter W. Wierwille, Characterization of Time-Varying Human Operator Dynamics (Project Icarus), NASA CR-539, August 1966.

31. Wierwille, Walter W., "A Theory and Method for Correlation Analysis of Nonstationary Signals," IEEE Trans., Vol. EC-14, No. 6, December 1965, pp. 909-919.

32. Gabay, Edward, and S. J. Merhav, "Identification of a Parametric Model of the Human Operator in Closed-Loop Control Tasks," IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-7, No. 4, April 1977, p. 284.

33. Denery, Dallas G., "An Identification Algorithm that is Insensitive to Initial Parameter Estimates," AIAA Journal, Vol. 9, No. 3, March 1971, pp. 371-377.

34. Allen, R. Wade, and Henry R. Jex, "A Simple Fourier Analysis Technique for Measuring the Dynamic Response of Manual Control Systems," IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-2, No. 5, November 1972, pp. 638-645.

35. Jex, H. R., and R. W. Allen, "Research on a New Human Dynamic Response Test Battery. Part I: Test Development and Validation. Part II: Psychophsiological Correlates," Proceedings of the Sixth Annual Conference on Manual Control, Air Force Institute of Technology/Air Force Flight Dynamics Laboratory, 7-9 April 1970, pp. 743-777.

36. Peters, Richard A., and R. Wade Allen, Operation Manual. Describing Function Analyzer. Model 1003 Serial 1001, Systems Technology, Inc., Working Paper No. 406-2, October 1970.

37. Magdaleno, Raymond E., "Serial Segments Method for Measuring Remnant," IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-2, No. 5, November 1972, pp. 674-678.

38. Hoh, Roger H., Thomas T. Myers, Irving L. Ashkenas, and Samuel J. Craig, Development of Handling Quality Criteria for Aircraft with Independent Control of Six Degrees of Freedom. Volume I: Analysis and Flight Test, Systems Technology, Inc., Technical Report No. 1135-1, June 1980.

39. Godfrey, K. R., "The Application of Pseudo Random Sequences to Industrial Processes and Nuclear Power Plant," Second IFAC Symposium, Identification and Process Parameter Estimation, Prague, Paper 7.1, 1970.

40. Aoki, M., and R. M. Staley, "On Input Signal Synthesis in Parameter Identification," Proceedings of the Fourth IFAC Congress, Warsaw, Paper 26.2. Also Automatica, 6, 1969, pp. 431-440.

41. Wingrove, Rodney C., Comparison of Methods for Identifying Pilot Describing Functions from Closed-Loop Operating Records, NASA TN D-6235, March 1971.

42. Ljung, Lennart, Ivar Gustavsson, and Torsten Soderstrom, "Identification of Linear Multivariable Systems Operating Under Linear Feedback Control," IEEE Trans. on Automatic Control, Vol. AC-19, No. 6, December 1974, pp. 8.

43. Gupta, Narendra K., W. Earl Hall, Jr., and Thomas L. Trankle, "Advanced Methods of Model Structure Determination from Test Data," Journal of Guidance and Control, Vol. 1, No. 3, May-June 1978, pp. 197.

44. Heffley, Robert K., and Wayne F. Jewell, Development of a CTOL Piloting Technique Measurement Scheme for a Real-Time Simulator Environment, NASA CR-152294, July 1979.

45. Wingrove, Rodney C., and Frederick G. Edwards, A Technique for Identifying Pilot Describing Functions from Routine Flight-Test Records, NASA TN D-5127, 1969.

46. Heffley, Robert K., Closed Loop Analysis of Manual Flare and Landing, AIAA Paper No. 74-834, August 1974.

47. Weir, D. H., and A. V. Phatak, Model of Human Operator Response to Step Transitions in Controlled Element Dynamics, NASA CR-671, January 1967.

48. Weir, David H., and Walter A. Johnson, Pilot Dynamic Response to Sudden Flight Control System Failures and Implications for Design, NASA CR-1087, June 1968.

49. Sadoff, Melvin, A Study of a Pilot's Ability to Control During Simulated Stability Augmentation System Failures, NASA TN D-1552, November 1962.

50. Elkind, J. I., J. A. Kelly, and R. A. Payne, "Adaptive Control Characteristics of the Human Operator in Systems Having Complex Dynamics," IEEE Proceedings of the Fifth National Symposium on Human Factors in Electronics, May 1964.

51. Clement, W. F., et al., Application of Manual Control Display Theory to the Development of Flight Director Systems for STOL Aircraft. Part II: Multi-Axis Sampling, Pilot Workload, and Display Integration, Systems Technology, Inc., Technical Report No. 1011-2, May 1972, Revised January 1974.

52. Hofmann, L. G., et al., "A New Link Value Estimator for Scanning Workload," Proceedings of the Ninth Annual Conference on Manual Control, May 23-25, 1973, pp. 379-387.

53. Clement W. F., et al., "A Direct Procedure for Partitioning Scanning Workload in Control Display System Design," Proceedings of the Ninth Annual Conference on Manual Control, May 23-25, 1973, pp. 389-400; also Proceedings of the IEEE 1973 International Conference on Systems, Man, and Cybernetics, November 1973, pp. 38-43.

54. Allen, R. W, W. F. Clement, and H. R. Jex, Research on Display Scanning, Sampling, and Reconstruction Using Separate Main and Secondary Tracking Tasks, NASA CR-1569, July 1970.

55. Clement, Warren F., R. Wade Allen, and Dunstan Graham, Pilot Experiments for a Theory of Integrated Display Format, JANAIR Report 711107, October 1971.

56. Enstrom, Kenneth D., and William B. Rouse, "Real-Time Determination of How a Human Has Allocated His Attention Between Control and Monitoring Tasks," IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-7, No. 3, March 1977, pp. 153-161.

57. Jewell, Wayne F., Data Collection Requirements, A Pilot Control Strategy Measurement Technique, and Verification Procedures for the Phase III HUD Program, Systems Technology, Inc., Technical Report No. 1153-1, February 1980.

58. Gai, E., and R. E. Curry, "Failure Detection by Pilots During Automatic Landing: Models and Experiments, Journal of Aircraft, Vol. 14, No. 2, February 1977, pp. 135-141.

59. Gai, E., and R. E. Curry, "A Model of the Human Observer in Failure Detection Tasks," IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-6, No. 2, February 1976, pp. 85-94.

60. Kleinman, David L., and Renwick E. Curry, "Some New Control Theoretic Models for Human Operator Display Monitoring," IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-7, No. 11, November 1977, pp. 778-784.

61. Ephrath, Ayre R., and Renwick E. Curry, "Detection by Pilots of System Failures During Instrument Landings, IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-7, No. 12, December 1977, pp. 841-848.

62. Spyker, D. A., S. P. Stackhouse, A. S. Khalafalla, and R. C. McLane, Development of Techniques for Measuring Pilot Workload, NASA CR-1888, November 1971.

63. Knemeyer, S., and G. L. Yingling, Results of the USAF Pilot Factors Program as They Apply to Flight Safety, Presented at Nineteenth Annual International Air Safety Seminar, 15-18 November 1966, Madrid, Spain; also presented at the Twelfth Annual Business Aircraft Safety Seminar, 15-18 May 1967, Washington, D. C.

64. von Wieser, M. F., and F. M. Wilson, Head-Up Display, Experimental Flight Test Programs, 14th Corporate Aircraft Safety Seminar, Flight Safety Foundation Inc., Washington, D. C., May 14, 1969.

65. Perry, D. H., W. G. A. Port, and J. C. Morrall, A Flight Study of the Sidestep Manoeuvre During Landing, Aeronautical Research Council, R and M 3347, July 1961.

66. Gartner, Walter B., and Miles R. Murphy, Pilot Workload and Fatigue: A Critical Survey of Concepts and Assessment Techniques, NASA TN D-8365, November 1976.

67. Schiflett, S. C., Operator Workload, An Annotated Bibliography, NATC Report No. SY-257R-76, 30 December 1976.

68. Anon., Methods to Assess Workload, AGARD-CP-216, April 1977.

69. Clement, Warren F., Annotated Bibliography of Procedures Which Assess Primary Task Performance in Some Manner as the Basic Element of a Workload Measurement Procedure, Systems Technology, Inc., Technical Report No. 1104-2, January 1978.

70. Levine, J. M., et al., Measurement of Workload by Secondary Tasks, A Review and Annotated Bibliography, ARRO Report R78-2, January 1978.

71. Anon., Assessing Pilot Workload, AGARD-AG-233, February 1978.

72. Wierwille, W. W., and R. C. Williges, Survey and Analysis of Operator Workload Assessment Techniques, Systemetrics Report S-78-101, September 1978.

73. Moray, N., (Ed.), Mental Workload, Its Theory and Measurement, Plenum Press, New York, 1979.

74. Jex, H. R., and W. F. Clement, "Defining and Measuring Perceptual Motor Workload in Manual Control Tasks," in N. Moray (Ed.), Mental Workload, Its Theory and Measurement, Plenum Press, New York, 1979, pp. 125-177.

75. Duning, Kenneth E., Craig W. Hickok, Kenneth C. Emerson, and Warren F. Clement, Control-Display Testing Requirements Study, AFFDL-TR-72-122, December 1972.

76. Heffley, Robert K., Robert L. Stapleford, and Robert C. Rumold, Airworthiness Criteria Development for Powered-Lift Aircraft. A Program Summary, NASA CR-2791 (Also FAA-RD-76-195), February 1977.

77. Harper, R. P., Jr., and George E. Cooper, A Revised Pilot Rating Scale for the Evaluation of Handling Qualities, Cornell Aeronautics Laboratories, Report No. 153, September 1966.

78. McDonnell, John D., Pilot Rating Techniques for the Estimation and Evaluation of Handling Qualities, AFFDL-TR-68-76, December 1968.

79. Cooper, G. E., "Understanding and Interpreting Pilot Opinion," Aeronautical Engineering Review, Vol. 16, No. 3, March 1957, pp. 47-52.

80. Kelley, C. R., and M. T. Wargo, "Cross-Adaptive Operator Loading Tasks," Human Factors, Vol. 9, No. 5, 1967, pp. 395-404.

81. Hudson, E. M., An Adaptive Tracking Simulator, Presented at the 1962 IRE International Congress on HFE, Long Beach, California.

82. Clement, Warren F., Investigation of the Use of an Electronic Multifunction Display and an Electromechanical Horizontal Situation Indicator for Guidance and Control of Powered-Lift Short-Haul Aircraft, NASA CR-137922, August 1976.

83. Merhav, S. J., and O. Ben Ya'acov, "Control Augmentation and Workload Reduction by Kinesthetic Information From the Manipulator," IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-6, No. 12, December 1976, pp. 825-835.

84. Jex, H. R., J. D. McDonnell, and A. V. Phatak, A Critical Tracking Task for Man-Machine Research Related to the Operator's Effective Delay Time. Part I: Theory and Experiments with a First-Order Divergent Controlled Element, NASA CR-616, October 1966.

85. McDonnell, J. D., and H. R. Jex, A Critical Tracking Task for Man-Machine Research Related to the Operator's Effective Delay Time. Part II: Experimental Effects of System Input Spectra, Control Stick Stiffness, and Controlled Element Order, NASA CR-674, January 1967.

86. Jex, H. R., J. D. McDonnell, and A. V. Phatak, "A Critical Tracking Task for Manual Control Research," IEEE Trans., Vol. HFE-7, No. 4, December 1966, pp. 138-145.

87. Jex, H. R., "Two Applications of the Critical Instability Task to Secondary Work Load Research," IEEE Trans., Vol. 8, No. 4, December 1967, pp. 279-282.

88. Ephrath, Arye R., A Novel Approach to the Cross-Adaptive Auxiliary Task, NASA TM X-73,170, May 1976, pp. 63-71.

89. Senders, J. W., Tracking with Intermittently Illuminated Displays, WADC TR-55-378, October 1955.

90. McGehee, W., Comparative Study of Pilot Fatigue Resulting From Extended Instrument Flight Using the Standard AAF and British Instrument Panels, Final Report, Project TED ATLR-601, US Navy, 1944.

91. Jones, R. E., J. L. Milton, and P. M. Fitts, Eye Fixations of Aircraft Pilots. Volume I: A Review of Prior Eye Movement Studies and a Description of a Technique for Recording the Frequency, Duration, and Sequences of Eye Fixations During Instrument Flight, USAF AMC TR No. 5837, September 1949.

92. Milton, J. L., R. E. Jones, and P. M. Fitts, Eye Fixations of Aircraft Pilots. Volume II: Frequency, Duration, and Sequence of Fixations When Flying the USAF Instrument Low Approach System (ILAS), USAF AMC TR No. 5839, October 1949.

93. Fitts, P. M., R. E. Jones, and J. L. Milton, "Eye Movements of Aircraft Pilots During Instrument-Landing Approaches," Aeronautical Engineering Review, Vol. 9, No. 2, February 1950, pp. 24-29.

94. Fitts, P. M., R. E. Jones, and J. L. Milton, Eye Fixations of Aircraft Pilots. Volume III: Frequency, Duration, and Sequence of Fixations When Flying Air Force Ground Control Approach System (GCA), USAF AMC TR No. 5967, February 1950.

95. Jones, R. E., J. L. Milton, and P. M. Fitts, Eye Fixations for Aircraft Pilots. Volume IV: Frequency, Duration, and Sequence of Fixations During Routine Instrument Flight, USAF AMC TR No. 5975, December 1949.

96. Milton, J. L., R. E. Jones, and P. M. Fitts, Eye Fixations of Aircraft Pilots. Volume V: Frequency, Duration, and Sequence of Fixations When Flying Selected Maneuvers During Instrument and Visual Conditions, USAF AMC TR No. 6018, August 1950.

97. Milton, J. L., B. B. McIntosh, and E. L. Cole, Eye Fixations of Aircraft Pilots. Volume VI: Fixations During Day and Night ILAS Approaches Using an Experimental Instrument Panel Arrangement, USAF TR No. 6570, October 1951.

98. Milton, J. L., B. B. McIntosh, and E. L. Cole, Eye Fixations of Aircraft Pilots. Volume VII: Fixations During Day and Night, GCA Approaches Using an Experimental Panel Arrangement, USAF TR No. 6709, February 1952.

99. Milton, J. L., and F. J. Wolfe, Eye Fixations of Aircraft Pilots. Volume VIII: Fixations During Zero-Reader Approaches in a Jet Aircraft, WADC TR 52-17, February 1952.

100. Cole, E. L., J. L. Milton, and B. B. McIntosh, Eye Fixations of Aircraft Pilots. Volume IX: Routine Maneuvers Under Day and Night Conditions, Using an Experimental Panel Arrangement, WADC TR 53-220, March 1954.

101. Watts, A. F. A., and H. C. Wiltshire, Investigation of Eye Movements of an Aircraft Pilot Under Blind Approach Conditions, The College of Aeronautics, Cranfield, College Note No. 26, May 1955.

102. Lennox, D., Airline Pilots' Eye Movements During Take-off and Landing in Visual Meteorological Conditions, Department of Supply, Australian Defence Scientific Service, Aeronautical Research Laboratories, Human Engineering Note 15, August 1963.

103. McRuer, D., H. R. Jex, W. F. Clement, and D. Graham, Development of a Systems Analysis Theory of Manual Control Displays, Systems Technology, Inc., Technical Report No. 163-1, October 1967.

104. Clement, W. F., D. T. McRuer, and R. H. Klein, Systematic Manual Control Display Design, presented at the 13th AGARD Guidance and Control Panel Symposium on Guidance and Control Displays, Paris, France, 19-21 October 1971.

105. Stackhouse, S. P., The Measurement of Pilot Workload in a Manual Control System, Honeywell, Inc., F0398 FR1, January 1976.

106. Gerathewohl, S. J., E. L. Brown, J. E. Burke, K. A. Kimball, W. F. Lowe, and S. P. Stackhouse, "Inflight Measurement of Pilot Workload: A Panel Discussion," Aviation, Space, and Environmental Medicine, Vol. 49, 1978, pp. 810-822.

107. Kalsbeek, J. W. H., and J. H. Ettema, "Physiological and Psychological Evaluation of Distraction Stress," Proceedings of the Second International Congress on Ergonomics, Taylor and Francis, Ltd., London, 1964, pp. 443-447.

108. Onishchenko, V. F., A. A. Volkov, and S. G. Mel'nik, "Application of the Poly-Effector Method for Evaluating the Psychophysiological Abilities of an Operator in the System Man-Machine," Problems of Engineering Psychology, NASA TT F-312, May 1965.

109. Kutayev-Smyk, L.A., I. P. Neumyvakin, and V. A. Ponomarenko, "Contribution to the Question of the Methods of Evaluating the Psychophysiological State of a Pilot in an Emergency Situation During Flight," Problems of Engineering Psychology, NASA TT F-312, May 1965.

110. Kalsbeek, J. W. H., "On the Measurement of Deterioration of Performance Caused by Distraction Stress," Ergonomics, Vol. 7, No. 2, 1964, pp. 187-195.

111. Schouten, J. F., J. W. H. Kalsbeek, and F. F. Leopold, "On the Evaluation of Perceptual and Mental Load," Ergonomics, Vol. 5, 1962, pp. 251-260.

112. Benson, Alan J., J. H. F. Huddleston, and John M. Rolfe, "A Psychophysiological Study of Compensatory Tracking on a Digital Display," Human Factors, Vol. 7, No. 5, October 1965, pp. 457-472.

113. Roman, James, "Long-Range Program to Develop Medical Monitoring in Flight -- The Flight Research Program. Volume I," Aero. Med., Vol. 36, No. 6, June 1965, pp. 514-518, et seq.

114. Anonymous, Report of Department of Transportation Air Traffic Control Advisory Committee, Volume 1, Department of Transportation, December 1969.

115. Levison, W. H., and J. I. Elkind, Studies of Multivariable Manual Control Systems; Two-Axis Compensatory Systems with Separated Displays and Controls, NASA CR-875, October 1967.

116. Hopkin, V. D., "Peripheral Vision and Flight Information," Human Problems of Supersonic and Hypersonic Flight, Proceedings of the Fifth European Congress of Aviation Medicine, London, 28 August - 2 September 1960.

117. Winblade, R. L., Current Research on Advanced Cockpit Display Systems, NATA, AGARD Report 491, October 1964.

118. Senders, J. W., "The Estimation of Pilot Workload," AGARD Ad Hoc Panel in Guidance and Control Symposium on "The Human Operator and Aircraft and Missile Control", Paris, September 5-6, 1966.

119. Weir, David H., and Richard H. Klein, The Measurement and Analysis of Pilot Scanning and Control Behavior During Simulated Instrument Approaches, NASA CR-1535, June 1970.

120. Spady, Amos A., Jr., and Max C. Kurbjun, "Flight Management Research Utilizing an Oculometer," SAE Paper No. 780527, International Air Transportation Meeting, Boston, Massachusetts, May 1-4, 1978.

121. Niemann, R. A., and K. Ziedman, Data Analysis for Driver Performance Studies. Volume II: A Computer Program for Analysis of Eye Movement Patterns, Amex Systems, Inc., Report AMEX-0105, November 1974.

122. Wolkovitch, J., R. E. Magdaleno, D. T. McRuer, D. Graham, and J. D. McDonnell, Performance Criteria for Linear Constant-Coefficient Systems with Deterministic Inputs, ASD-TR-61-501, Dec. 1961.

123. Magdaleno, R. E., and J. Wolkovitch, Performance Criteria for Linear Constant-Coefficient Systems with Random Inputs, ASD-TDR-62-470, Jan. 1963.

# APPENDIX A

## SECTION IV FROM STI TR-1156-1

## CLASSIFICATION OF THE SOURCES AND DISTINGUISHING CHARACTERISTICS OF ERROR

## SECTION IV

## CLASSIFICATION OF THE SOURCES AND DISTINGUISHING CHARACTERISTICS OF ERROR

A thorough evaluation of piloting and traffic controlling tasks among mission phases within the national airspace environment is a prerequisite for planning research on or conducting an investigation of human error which employs full mission simulation. The importance of this prerequisite has been emphasized by the example of the approach and landing tasks at the end of Section III. Having thus identified at least some of the potential for human error among normal operations, we turn our attention in this section to the abnormal — classification of the sources and distinguishing characteristics of error itself.

Another prerequisite for planning and conducting research in any discipline is a set of accepted definitions. For example, such terms as defect, failure, reliability, unscheduled maintenance, and performance measurement have acquired disciplined meaning where applied to purely machinelike systems. An analogous glossary of terms is not yet widely accepted for analysis of human reliability and performance. In the next topic, therefore, we shall adopt several definitions of error already proposed and qualify the meaning of others.

### A. DEFINITIONS OF ERROR

As we have already remarked, errors or mismatches between desired and actual system or subsystem outputs are the sine qua non of situations where feedback is involved as an operating principle. Most of the time human operators use these errors to advantage in performing as error-correcting rather than error-avoiding system elements. For this reason in operations involving pilots, air crew, and ATC, the errors per se are of major concern only when they are undesirable because of their size, timing, or character. These errors, which are intolerable in one way or another, we shall call grievous errors.

In general, a grievous error will involve an exceedence of safe operating tolerances. "System error" and "system deviation," terms used by the FAA Air Traffic Control Service to describe procedural errors, missed acquisitions, and extreme deviations that lead to interactions between two aircraft, are grievous errors. These may derive from malfunctions or failures of system components which result in degraded system operation. Alternatively they may stem from the impact on a normally operating system of an unexpectedly severe forcing function or disturbance. This is an instance of what Singleton (Ref. 41) refers to as a substantive error, non-intended performance because the problem was inadequately defined at the outset, before the system requirements and specifications were established, or the system design itself was inadequate.

Singleton also introduces the term formal error to apply to cases where some rule has been broken. Grievous errors in general can be verified quantitatively because exceedences of tolerances can usually be measured. On the other hand, transgressions of a rule may not necessarily be observable or measurable, unless the rule specifies a commensurate tolerance. Out-of-sequence performance (within tolerances otherwise) is an example of transgression of a rule which might very likely be observable.

The substantive and formal error classifications are useful in setting up a taxonomy of human error definitions. In general human error ≡ inconsistency with a predetermined behavioral pattern used in establishing system requirements, specifications, and the resulting design (Ref. 42) and in defining the procedures to be used as well. Then,

1) Formal (human) error ≡ transgression of a rule, regulation, algorithm (Refs. 41 and 43), or out-of-sequence performance (Ref. 44).

2) Incoherent (human) error ≡ non-required performance, i.e., output not stimulated by an input (Ref. 44).

3) Substantive (human) error ≡ non-intended performance, e.g., because the procedure was inadequately defined.

Human errors that do not always result in grievous errors may be nearly impossible to measure in practice unless behavioral identification techniques

are employed. Behavioral identification may be performed by qualified observers (Refs. 24, 45, and 46) or by signal correlation analysis which can partition human error into coherent and incoherent components. Such identification of human errors which may be inconspicuous in one situation is very important, for they may lead to grievous errors in other circumstances.

## B. SOURCES AND CAUSES OF HUMAN ERROR

The functional pathway triad and metacontroller model for human behavior developed in Section III contains within its structure many features which can, in abnormal versions, lead to grievous system errors. These features we shall refer to as sources or antecedents of error. Sources are endogenous or internal to the human. Their consequences are all measurable in terms of changes from ideal or nominal human behavior for a particular task. These changes may be induced by external (exogenous) factors which will be referred to as causes of error. The first two columns of Table 9 illustrate these distinctions for compensatory operations.

The remaining two columns of Table 9 present a verbal synthesis of a great deal of empirical data from many experimenters. All of the currently demonstrated forms of abnormal compensatory input-output behavior are represented here. In total they represent an error source which can be described generally as

> inappropriate perception, decision, and/or execution within a selected level (in this case, compensatory) of organization of behavior.

The sources of error in this framework are summarized in Table 10.

In principle tables similar to Table 9 can be constructed for the other source possibilities in Table 10, e.g., Table 11 for pursuit operations. However the experimental data base for most of these is nowhere near as comprehensive as it is for the compensatory pathway. Many of the elements in the precognitive pathway can be developed, by analogy, from Table 1 of Ref. 32, which lists the presumed sources of "slips" (or errors) in the structure of Fig. 10b.

# TABLE 9

## BEHAVIORAL SOURCES OF ERROR IN COMPENSATORY SYSTEMS

SINGLE CHANNEL OPERATIONS

| BASIC SOURCE (ENDOGENOUS) | CAUSES (EXOGENOUS) | OPERATOR BEHAVIOR | EFFECTS ON SYSTEM |
|---|---|---|---|
| Extreme command or disturbance amplitudes | Unexpectedly large command or extreme environment | Operator response normal | System overloaded, forced out of tolerance although operating properly |
| Extreme command or disturbance bandwidth | Broadband input signal noise; Unexpectedly broadband disturbance | Regression of crossover frequency | Reduced system bandwidth |
| Controlled-element change | Mulfunction/failure in controlled element | Affecting output for transient interval; Adaptation to new controlled element | Transient errors during transition; Reduced system bandwidth |
| Reduced attention field | Poor signal/noise ratio (e.g., poor contrast, high intensity distraction stimuli, low level signals, etc.) | Operator threshold, net gain reduction | System bandwidth reduction; (missed signals as one extreme) |
| Reversals | Misperception of error sign; Naivete | Remnant increase; Intermittently reversed output | Increased system noise; Intermittently reversed system output |

MULTI-INPUT OPERATIONS

| BASIC SOURCE (ENDOGENOUS) | CAUSES (EXOGENOUS) | OPERATOR BEHAVIOR | EFFECTS ON SYSTEM |
|---|---|---|---|
| Divided attention, perceptual scanning | Increased informational requirements for monitoring or control | Remnant increase (scanning); Increase in loop gains; Simultaneous multi-channel operations | Increased system noise; Reduced bandwidth |
| | Information overload: Too many separate input channels; Too many significant signals; Backlog of unattended operations | As above, plus failure to detect some signals, increased latencies, and missed output responses | Saturation; Missed responses; Instability in the mean square sense |
| Reduced attentional field | Operator impairment (fatigue, alcohol, hypoxia, etc.) | Remnant increase over scanning; Further decrease in loop gain; Sequentially-switched single channel operations; Deletion/missed responses | Increased system noise Reduced bandwidths Increased latencies Missed responses |
| Illusions, kinetosis | Conflict between or among visual, vestibular, aural, kinesthetic and/or proprioceptive inputs | Remnant increase; Decrease in operator's gain; Mal à propos responses; Missed responses | Increased system noise Reduced bandwidth Mal à propos responses Missed responses |

TABLE 10

SOURCES OF HUMAN ERROR

(Sources are endogenous or internal to the human operator by definition)

- Inappropriate perception, decision, and/or execution within a selected level of behavioral organization
  - Compensatory (expanded in Table 9)
  - Pursuit (expanded in Table 11)
  - Precognitive (expanded in Table 1 of Ref. 32)
    - Selection of response unit
    - Execution of response
- Transitions from a higher to lower level of behavioral organization
  - Precognitive to pursuit
  - Precognitive to compensatory
  - Pursuit to compensatory
- Inappropriate organization of perception and behavior for the task at the executive level of the metacontroller
  - (Expanded in Table 12 for the cockpit environment)
  - (Expanded in Table 13 for the traffic control environment)
- Inadequate off-line monitor/supervisor in the metacontroller

TABLE 11

BEHAVIORAL SOURCES OF ERROR IN PURSUIT OPERATIONS

(Multi-Input Operations, by Definition)

| BASIC SOURCE (ENDOGENOUS) | CAUSES (EXOGENOUS) | OPERATOR BEHAVIOR | EFFECTS ON SYSTEM |
|---|---|---|---|
| Controlled element change | (see corresponding causes in Table 9) | Transient regression to compensatory level (see corresponding behavior in Table 9) | Transient errors during transition;<br>Reduced system bandwidth |
| Divided attention, perceptual scanning | (see corresponding causes in Table 9) | Remnant increase;<br>Decrease in operator's gain;<br>(see also corresponding behavior in Table 9) | Increased system noise;<br>Reduced bandwidth;<br>(see also corresponding effects in Table 9) |
| Reduced attentional field in spatial dimensions | Poor input and/or error signal/noise ratio (e.g., inability to identify input.)<br>Task involves disturbance regulation rather than command-following and disturbance cannot be identified;<br>Mismatched scaling between input and error;<br>Distortion of input;<br>Lack of input conformability with visual field;<br>See also corresponding causes in Table 9 | Remnant increase;<br>Operator's threshold on input may cause missed responses and regression to compensatory level;<br>Operator's threshold on error may reduce gain in or open compensatory loop<br>(see also corresponding behavior in Table 9) | Increased system noise;<br>Reduced system bandwidth (missed responses as one extreme) |
| Reduced attentional field in temporal dimension, i.e., reduced preview | Inability to identify <u>future</u> input or disturbance;<br>Prodigious extrapolation required to estimate <u>future</u> input or disturbance | As above, plus increased latencies | As above, plus increased response latencies |
| Reversals | Perceptual inversion of input;<br>Faulty input-background discrimination;<br>Lack of input conformability with visual field | Remnant increase;<br>Intermittently reversed output | Increased system noise;<br>Intermittently reversed output |
| Illusions, kinetosis | (see corresponding causes in Table 9) | Remnant increase;<br>Decrease in operator's gain;<br>Mal à propos responses;<br>Missed responses | Increased system noise;<br>Reduced bandwidth;<br>Mal à propos responses;<br>Missed responses |

TABLE 12

CAUSES OF ERROR LEADING TO INAPPROPRIATE ORGANIZATION OF PERCEPTION AND BEHAVIOR AT THE EXECUTIVE LEVEL OF THE METACONTROLLER IN THE COCKPIT ENVIRONMENT

Items 1-5 are associated with the "situation identification" block in Fig. 10a

Item 6 is associated with the "selection of appropriate pathway(s)" in Fig. 10a

Errors in:

(1) Formulation of intent, assignment of function (to crew member by captain) and its priority

- Tactical Decisions (assignment retained by captain with rare exceptions)
- CNI
- Systems Operation
- Flight Control

(2) Identification of specific task/situation/action: <u>continuous</u> or <u>discrete</u>

- Information retrieval (e.g., checklists, clearance, instructions, manuals, maps, SIDs, STARs, approach plates)
- Conferring to arrive at a decision
- Monitoring
- Controlling/commanding
- Command-Interpretation and transcription (e.g., clearance, etc.)
- Command-following (e.g., flying)
- Disturbance regulation
- Deferring action (changing priority)
- Reassignment of action (to a different crew member by captain)

(3a) Selection of likely sources of information and their temporal order (i.e., stale, current, or preview)

- Checklists, clearances, instructions, manuals, maps, SIDs, STARs, approach plates
- Voice advisory or command
- Visual field
- Relevant instruments/displays/annunicators
- Motion cues
- Proprioceptive cues

(continued on next page)

TABLE 12 (Concluded)

Errors in:

(3b) Assignment of priority in sources of information among inputs, feedbacks

Specific IFR sources
Specific VFR sources
Type of display: compensatory, pursuit, preview

(4) Identifying predictability or coherence in and among sources of information

Patterns in random commands, disturbances - nil
Patterns in wind shears - may be highly correlated
Patterns in programmed commands, maneuvers
Patterns in periodic commands, disturbances
Patterns in discrete commands, disturbances, failures
Patterns in slowly divergent or ramp-like disturbances, failures

(5) Identifying familiarity with task

Nil
Slight
Moderate
Great, i.e., very well rehearsed

(6) Organizing operation on inputs, feedbacks:

Continuous or discrete operations
SOP level: compensatory, pursuit, precognitive, combinations
Loop structure
Behavioral adaptation within loop structure
Specific cued (behavioral) programs

TABLE 13

CAUSES OF ERROR LEADING TO INAPPROPRIATE ORGANIZATION OF PERCEPTION AND BEHAVIOR AT THE EXECUTIVE LEVEL OF THE METACONTROLLER IN THE TRAFFIC CONTROL ENVIRONMENT

Items 1-5 are associated with the "situation identification" block in Fig. 10a

Item 6 is associated with the "selection of appropriate pathway(s)" in Fig. 10a

Errors in:

(1) Formulation of intent, assignment of function (to specialist by supervisor) and its priority

ATC: Enroute, terminal (departure, approach), final, surface

Commercial: Aircraft dispatcher, ramp control supervisor, area operations supervisor, operations controller

(2) Identification of specific task/situation/action: <u>continuous</u> or <u>discrete</u>

Information retrieval
Communication input
Conferring to arrive at a decision
Surveillance, searching, pattern recognition
Monitoring
Tracking
Controlling/commanding/advising/interrogating (communication output)
Deferring action
Reassignment of action (to a different specialist)

(3a) Selection of likely sources of information and their temporal order (i.e., stale, current, or preview)

Visual: Flight progress posting strips/ETABS
PPI/ATCRBS/DABS

Aural communications

(3b) Assignment of priority in sources of information among inputs, feedbacks

Specific visual sources
Specific aural sources
Type of display: compensatory, pursuit, preview

(continued on next page)

TABLE 13 (Concluded)

Errors in:

(4) Identifying predictability or coherence in and among sources of information

- Patterns in programmed tracks on PPI
- Patterns in predicted courses on PPI
- Patterns in programmed altitude responses
- Patterns in predicted altitude responses
- Patterns in overall flight progress
- Patterns in discrete commands, disturbances, failures
- Patterns in slowly divergent or ramp-like disturbances, failures
- Coherence in aural communications
- Interference in aural communications

(5) Identifying familiarity with task

- Nil
- Slight
- Moderate
- Great, i.e., very well rehearsed

(6) Organizing operation on inputs, feedbacks

- Continuous or discrete operations
- SOP level: compensatory, pursuit, precognitive, combinations
- Loop structure
- Behavioral adaptation within loop structure
- Specific cued (behavioral) programs (e.g., conflict alert and collision avoidance command)

Transitions from higher to lower levels occur when the attentional field becomes too narrow. They can also occur when the human is sufficiently impaired perceptually (i.e., by alcohol, fatigue, hypoxia, etc.) so that action as a multi-channel operator is significantly degraded. In these instances divided attention is possible only by switching to and fro as an essentially single channel information processing device.

Although probably one of the most fundamental sources of human error, the inappropriate organization of perception and behavior for the task at the executive level of the metacontroller has received much less attention in the literature than have inappropriate perception, decision, and/or execution within a selected level of behavioral organization. The SOP theory described in Section III offers a unifying approach to inappropriate organization as a source of human error. To illustrate this source more specifically, we have partitioned possible causes of error leading to inappropriate organization of perception and behavior in two contexts, the cockpit environment and the traffic control environment. (There are actually two traffic control environments, one operated by the Federal Aviation Administration, the other, peculiar to each commercial operator. For the purpose of classifying these causes of error among traffic controllers, however, one list will suffice; the other list will serve the cockpit.) Table 12 presents the partition for the cockpit, and Table 13, for the traffic control environment. Within each subdivision, specific examples are listed to help in understanding the meaning of the subdivision.

This concludes our subdivision of the causes of error. Next we shall consider the assignment of causes and some remedial actions.

## C. ATTRIBUTION OF ERROR (ASSIGNMENT OF CAUSE OR RESPONSIBILITY FOR ERROR)

Singleton, in Ref. 41, identifies significant problems in addressing scientifically the issue of assigning responsibility for error.

> "Most societies have not resolved the distinction between two main approaches (to attribution). One assumes that human beings are responsible for their own actions and are therefore responsible for the errors they make. The opposite

> view is that errors are an inherent component in all human performance, that they should be planned for and designed for and when they do occur the fault should be traced to the system designer rather than the operator. At the individual level, few people are sufficiently self-confident to deliberately acknowledge their own mistakes, particularly if there are financial consequences in doing so. This is an especially difficult problem in the insurance world, where accidents are investigated with a view to deciding who is going to pay for the damage caused either to people or to property. In such a situation it is not surprising to find that it is impossible to regard the evidence as scientific in any sense."

One of the prime justifications for the study of full mission operations in the Man Vehicle Systems Research Facility is to avoid these problems gracefully. Another way is to sidestep the issue of attribution in order to acquire incipient and consummate error data with a semblance of scientific credibility. The NASA Aviation Safety Reporting System (Ref. 47) is a prime example of a confidential, non-punitive program designed to sidestep the issue of attribution in the process of acquiring a scientifically useful error data base.

Notwithstanding the aforementioned problems, we believe that there may be useful ways to classify the assignment of causes of error in an impersonal way which has scientific value. Such a classification is presented in Table 14. The subdivisions of attribution shown there were selected so that they could be identified with constructive remedial action. Examples of such remedies are listed on the right hand side of the table. Some of these, e.g., skill development and continuing rehearsal for proficiency maintenance, have been discussed thoroughly in Sections II and III.

# TABLE 14

## PARTITIONS OF ATTRIBUTION AND REMEDY

| ATTRIBUTION<br>Assignment of Causes of Error* | REMEDY<br>Correction of Cause |
|---|---|
| • Inadequate definition of the problem at the outset before the system requirements and specifications were established. (Produces substantive or existential error, because the system specification itself is inadequate) otherwise called "unforeseen circumstances." | • Design modification |
| • Inadequate system design (presumes the specifications are adequate, but their interpretation in terms of the design is not adequate; therefore also produces substantive error.) | • Design modification |
| • Inadequate definition of the procedures (really part of system design, but emphasizes modus operandi and therefore also produces substantive error) | • Procedural modification |
| • Naiveté | • (Naiveté) |
| — Mismatched or misapplied skills. | — Selection and training for skill development |
| — Ignorance of regulations or rules | — Explanation and training |
| — Inadequate instruction of the procedures | — Retraining and rehearsal |
| • Inadequate interpretation and/or execution of the procedure(s) | • (Inadequate interpretation and/or execution of the procedure(s)) |
| — Lapse in practice | — Continuing rehearsal for proficiency maintenance |
| — Psychophysiological stressors | — (Psychophysiological stressors) |
| (1) Workload | (1) Redistribution of some functions or tasks among crew members or reassignment of some functions to automatic control |
| (2) Environmental disorders | (2) Correction or reassignment |
| (3) Emotional disorders | (3) Reassignment, rehabilitation |
| (4) Alcohol, drugs | (4) Reassignment, rehabilitation |
| — Psychoneurosis | — Reassignment, rehabilitation |
| — Blunders — everyone involved thinks that everything is okay when it isn't. | — Requires an independent observer or agency to monitor, recognize, and correct. |
| — External disturbances (i.e., external to the human operator), e.g., wind shear; potential traffic conflicts; failures of the machine or system | — Design modification to sense the disturbance, if possible, so that the operator can adopt pursuit or precognitive levels of behavior to cope with the disturbance where the compensatory level is inappropriate; design modification to improve reliability of the machine, possibly even by reassignment of some functions to a human operator not otherwise overloaded. |

* The absence of assignable cause means that the error will be called "chance" or "random."

## APPENDIX B

## ADAPTIVE PSYCHOMOTOR AND COGNITIVE TASKS FOR MEASURING EXCESS CONTROL CAPACITY

(From Refs. B-1, B-2, and B-3)

The considerable pilot rating data available in Ref. B-4 for the estimation of handling qualities indicate that, where closed-loop compensatory tracking is the task, the pilot's increments in rating are indeed based on the relative difficulty with which he obtains and maintains the specified performance. This notion that among the causal factors of pilot rating are the pilot's attempts to maintain performance by working to control in spite of the increasing difficulty was further supported by an experiment which measured a parameter uniquely related to excess control capacity (Ref. B-4).

A secondary tracking task* was used to "load" the pilot so that his performance on the primary task began to deteriorate. A block diagram of these tasks is shown in Fig. B-1. The difficulty of the secondary task was made proportional to primary task performance. Thus when the pilot was keeping primary task error performance less than a criterion value, E, the secondary task difficulty was automatically increased by increasing the rate of divergence of the secondary instability. Conversely, when the pilot was so busy with the secondary task that primary error was larger than the criterion value, the secondary task difficulty automatically decreased. The final stationary level of secondary difficulty was determined by the sensitivity of the primary task performance to loading. The final "score" is $\lambda_x$, the stationary value of the secondary unstable pole ($\lambda$) in rad/sec. The scores obtained from this cross-coupled secondary task represent its degree of difficulty;

---

* The adjective "subcritical" implies that $0 < \lambda < \lambda_c$, where $\lambda_c$ is the "critical" upper bound at which the human operator loses control of the secondary task instability with no primary task. $\lambda_c$ is a function of the operator's effective time delay in tracking, which is the analog of the operator's discrete reaction time delay or latency.

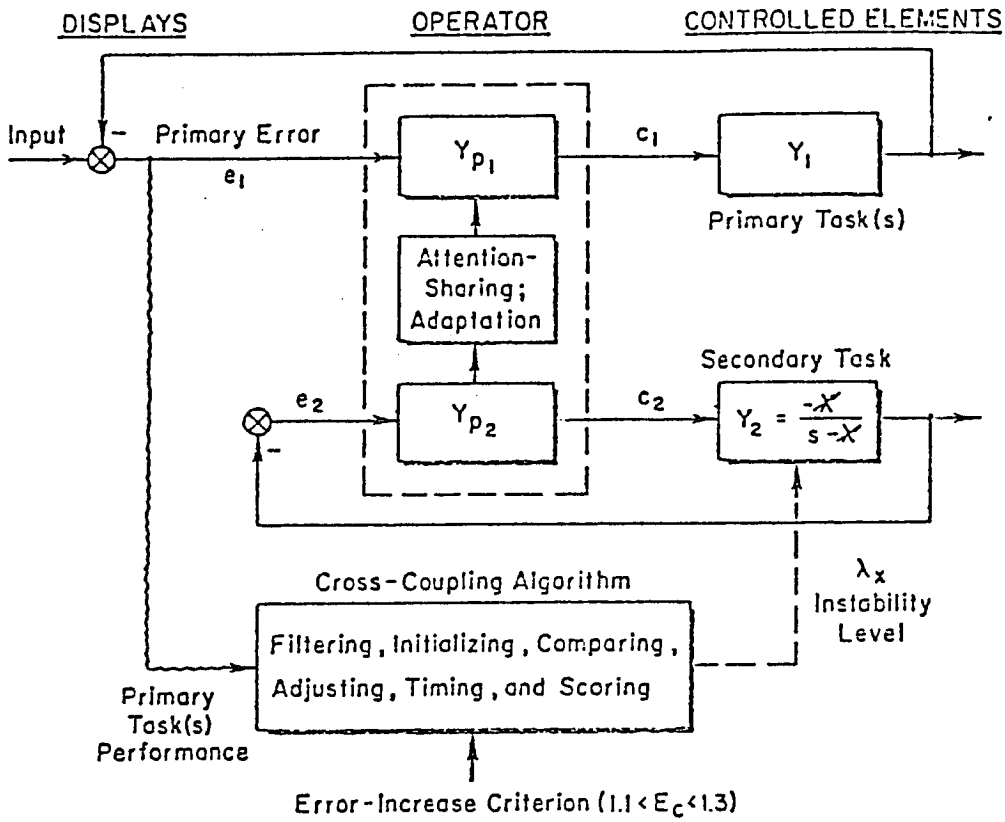



Figure B-1. Elements of the Cross-Coupled Instability Task (CCIT) (From Ref. B-5)

consequently, they also represent the "degree of ease" of the primary task or the excess control capacity available with respect to the primary task.

## ADAPTING THE CROSS-COUPLED SUBCRITICAL PSYCHOMOTOR TASK FOR A SPECIFIC CONTEXT

Referring again to Fig. B-1, notice that a given primary task or ensemble is monitored for task performance error, which is allowed by criterion E to grow not more than 10 to 30 percent over the unloaded performance error, measured at the beginning of each run to normalize effects of skill, learning, and individual variations from session to session. Special filtering and trend circuits detect when the unloaded

primary task performance error is stable, at which point the unloaded rms performance error ($\sigma_e$) is logged for later use, and the cross-coupling activated. A plausible secondary task in the operator's primary task context is simulated with a first-order-instability* whose level is slowly increased as long as the smoothed primary task error is less than the "error-increase criterion" (E ≡ loaded rms error/unloaded rms error, where $1.1 < E < 1.3$). As the actual primary task error ratio increase approaches E, the slow growth in the cross-coupled instability becomes asymptotic and its average is scored as the cross-coupled-limit, $\lambda_x$. The "Excess Control Capacity," EC (an index of workload margin) is found by dividing $\lambda_x$ by $\lambda_c$, the subjects' critical instability score for the same session, using the secondary task control and display with no primary task:

$$EC \equiv \lambda_x/\lambda_c \Big|_{\text{same Ss, session, task}} \qquad \text{(B-1)}$$

As previously established, $\lambda_x$ is an inverse measure of the fraction of time the operator can spend away from the primary task; thus it is a direct measure of excess control capacity. Normalizing by the individual concurrent level of $\lambda_c$ makes the EC score truly representative of workload margin and not just skill in secondary task tracking. Reference B-5 describes the development of this task, the detailed operation, and a series of experiments which validate the assumption that the primary task behavior is not changed in form and by only a small and controlled degree.

Individual measurements of excess control capacity for each of two or more primary tasks can be combined by a multiplication process (Ref. B-6) to estimate the combined value of EC which would be measured

---

* The adjustable first order instability can serve as a surrogate for either an integration or an instability in the equations describing the controlled element.

if all of the given "primary" tasks were performed in concert. The combined value of EC is given by the product of the individual values of EC:

$$(EC)_n = \prod^{n} (EC)_i \qquad \text{(B-2)}$$

This empirical "product rule" has been validated with more extensive multiaxis Cooper-Harper rating data in Ref. B-7. In effect, the product rule results in the physically satisfying vector addition of individual and combined fractional values of EC, regardless of the number of "primary" tasks.

For an overall figure of performance, we sometimes calculate a Performance Penalty index, P, which combines the input-normalized error with the inverse of excess control capacity (call it workload index $\equiv \lambda_c / \lambda_x$)

$$P \equiv \frac{\sigma_e}{\sigma_i} + \frac{\lambda_c}{\lambda_x} \qquad \text{(B-3)}$$

Where $P$ = Performance Penalty

$\sigma_e$ = rms unloaded error

$\sigma_i$ = rms input

$\lambda_c$ = critical instability with no primary task

$\lambda_x$ = cross-coupled instability

Since $\sigma_{e_{loaded}} / \sigma_{e_{unloaded}} < E$, the normalized error criterion, a better tracker can still achieve a lower penalty index P even if the workload index is comparable among Ss.

ADAPTING A CROSS-COUPLED COGNITIVE TASK
FOR A SPECIFIC CONTEXT

Another type of secondary task — this one discrete — has higher face validity in terms of cognitive monitoring, processing, and acknowledging an advisory message rather than performing continuous psychomotor activity for the purpose of control. The discrete secondary advisory stimulus can be communicated visually or aurally. (If visual, the advisory stimulus is usually outside the foveal field of the primary task display.)

The fundamental measure of the operator's reserve cognitive capacity with respect to the primary task is proportional to the operator's average response time latency (RT) to an ensemble of the secondary advisory stimuli. Various types of latency can be measured, e.g., simple reaction time (Refs. B-8 through B-11), disjunctive or choice reaction time (Refs. B-8, B-9, B-12, and B-13), or compound choice reaction time* (Refs. B-14 through B-17). The measure of excess cognitive capacity is usually interpreted as $(RT)_o/(RT)_L$, where $(RT)_o$ is the operator's average response latency to the secondary advisory stimuli while the operator is concentrating solely on the secondary task (i.e., not performing the primary task(s)) and $(RT)_L$ is the operator's average

---

* Commonly called the "Sternberg item recognition time." The Sternberg short-term memory task is an information processing task designed to assess cognitive reserve capacity under primary task loading conditions. The operator memorizes designated "critical" sets of N items where N is an integer > 0 (e.g., N specific letters, numbers, words, or symbolic characters) which are selected beforehand from a larger sample space. Items which are not members of the critical set are, by definition, "non-critical." A displayed item, chosen at random from the sample space is communicated to the operator visually or aurally to serve as the stimulus. The operator has to identify the item as "critical" or "non-critical" and provide the appropriate discrete response, usually by means of a two-way switch, within a prescribed time limit. Responses are recorded and evaluated in terms of latency and correctness. The average response time latency, $\overline{RT}(N)$, is a linear function of N. Increases in the slope of the Sternberg function [$\overline{RT}(N)$ versus N] are a measure of higher cognitive loads imposed by concurrent primary tasks. Increases in the extrapolated intercept of the Sternberg function as $N \rightarrow 0$ are a measure of higher perceptual motor loads imposed by concurrent primary tasks.

loaded response latency to the secondary advisory stimuli while the operator is performing the primary task(s).

Usually, although not necessarily, in the use of this type of secondary task, one presents a subsequent advisory stimulus to the operator as soon as the previous one is responded to. The operator is nevertheless instructed to regard a particular task or set of (other) tasks as "primary" and to respond to the designated secondary advisory stimulus only if the operator believes he can do so without compromising his performance on the primary task(s). The intent of this instruction, of course, is to minimize interference with or "loading" of the primary task. In practice, however, a definite loading of the primary task occurs. Such loading may be constrained and regulated by cross-coupling the average presentation or generation rate of the secondary advisory stimuli to a measure of primary task performance in the manner of Fig. B-2, which combines the methods of Refs. B-5 and B-18.

Figure B-2 is analogous to Fig. B-1, except for the difference in the type of secondary task and the fact that the cross-coupling signal, $\lambda_y$, is the average random character generation rate in Fig. B-2 instead of the instability level, $\lambda_x$, in Fig. B-1. The reciprocal of $\lambda_y$ ($1/\lambda_y$) is therefore the mean time between secondary task advisory stimuli. Consequently $1/\lambda_y$ subsumes the operator's average response time latency, RT, and includes any additional latency which is necessary to prevent loading the primary task beyond the error increase criterion, E. As the actual primary task error ratio increase approaches E, the slow growth in character generation rate, $\lambda_y$, becomes asymptotic and its average value is scored. The "EXcess Cognitive Capacity," XC (an index of workload margin), is found by dividing $\lambda_y$ by $\lambda_o$, the asymptotic value of $\lambda$ for the same session, using the secondary advisory task and display with no primary task:

$$XC \equiv \lambda_y/\lambda_o \Big|_{\text{same Ss, session, task}} \qquad \text{(B-4)}$$

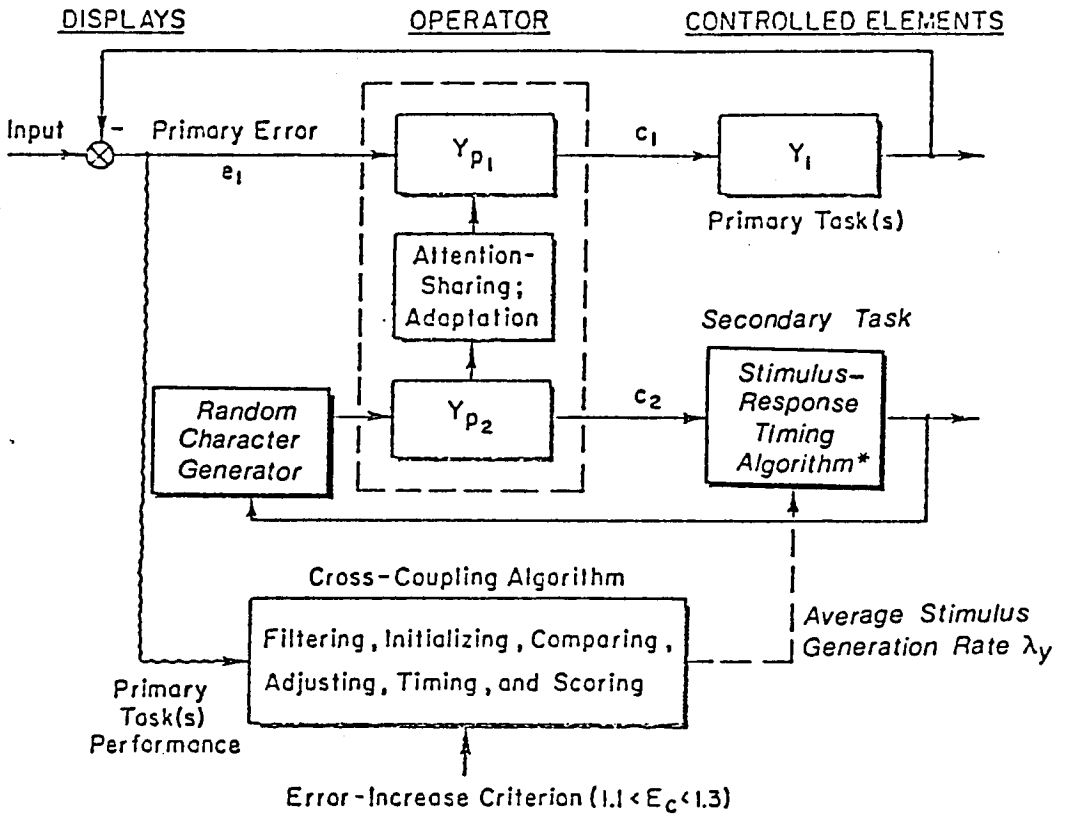



Figure B-2. A Cross-Coupled Adaptive Cognitive Task

* The algorithm for enabling the random character generator can be as follows:

Character generator initially off.

Compute $h = \lambda_y T_F$ where $\lambda_y$ is average stimulus generation rate from the cross-coupling algorithm, which includes average primary task error(s) and average secondary task response time. $T_F$ is computation frame time

Each computation frame generates (from a uniform probability distribution) a random number (x) with $0 \leq x \leq 1$ such that

if $0 \leq x \leq h$, enable the character generator
if $h < x \leq 1$, make no change in the state of the character generator; recompute h and recycle the test on h.

If the character generator is enabled, disable the above test on h and measure the time until the operator's response is received or until the time limit expires, whichever is less. Weight incorrect responses with a penalty proportional to the time limit.

When a response is made by the operator, disable the character generator.

If the response is correct, recompute h and recycle the test on h.

If the response is missed or is incorrect, wait until the time limit or penalized time limit expires before recomputing h and recycling the test on h.

Alternatively one may, as in Ref. B-18, calculate an overall performance measure which combines the unloaded primary input-normalized error with the excess cognitive capacity as the quotient,

$$Q = \left(\frac{\lambda_y}{\lambda_o}\right) \Big/ \left(\frac{\sigma_e}{\sigma_i}\right) \qquad \text{(B-5)}$$

Again, since $\sigma_{e_{unloaded}}/\sigma_{e_{unloaded}} \leq E$, the normalized error criterion, a lower primary error will be reflected in a higher Quotient, Q, even if the excess cognitive capacity is comparable among Ss.

## REFERENCES


B-1 Jex, H. R., and W. F. Clement, "Defining and Measuring Perceptual Motor Workload in Manual Control Tasks," in N. Moray (Ed.), Mental Workload, Its Theory and Measurement, Plenum Press, New York, 1979, pp. 125-177.

B-2 Clement, W. F., D. T. McRuer, and R. H. Klein, "Systematic Manual Control Display Design," Guidance and Control Displays, AGARD-CP-96, February 1972.

B-3 Duming, Kenneth E., Craig W. Hickok, Kenneth C. Emerson, and Warren F. Clement, Control-Display Testing Requirements Study, AFFDL-TR-72-122, December 1972.

B-4 McDonnell, J. D., Pilot Rating Techniques for the Estimation and Evaluation of Handling Qualities, AFFDL-TR-68-76, December 1968.

B-5 Jex, H. R., W. F. Jewell, and R. W. Allen, "Development of the Dual-Axis and Cross-Coupled Critical Tasks," Eighth Annual Conference on Manual Control, AFFDL-TR-72-92, January 1973, pp. 529-552.

B-6 Clement, Warren F., Lee Gregor Hofmann, and Richard E. Blodgett, Application of Manual Control Display Theory to the Development of Flight Director Systems for STOL Aircraft. Part II: Multi-Axis Sampling, Pilot Workload, and Display Integration, Systems Technology, Inc., Technical Report No. 1011-2, January 1974.

B-7 Ashkenas, Irving L., Roger H. Hoh, and Samuel J. Craig, Recommended Revisions to Selected Portions of MIL-F-8785B(ASG) and Background Data, AFFDL-TR-73-76, August 1973.

B-8 Woodworth, Robert S., and Harold Schlosbert, Experimental Psychology, Ref. ed., Holt, Rinehart and Winston, 1964.

B-9 McDonnell, J. D., and H. R. Jex, A "Critical" Tracking Task for Man-Machine Research Related to the Operator's Effective Delay Time. Part II: Experimental Effects of System Input Spectra, Control Stick Stiffness, and Controlled Element Order, NASA CR-674, January 1967.

B-10 Melanson, D., et al., "The Effect of Communications and Traffic Situation Displays on Pilots' Awareness of Traffic in the Terminal Area," Proceedings of the Ninth Annual Conference on Manual Control, NASA CR-142295, May 1973, pp. 25-39.

B-11 Clement, Warren F., Investigation of the Use of an Electronic Multifunction Display and an Electromechanical Horizontal Situation Indicator for Guidance and Control of Powered-Lift Short-Haul Aircraft, NASA CR-137922, August 1976.

B-12 Spyker, D. A., S. P. Stackhouse, A. S. Khalafalla, and R. C. McLane, Development of Techniques for Measuring Pilot Work Load, NASA CR-1888, 1970, p. 109.

B-13 Ephrath, Arye R., "A Novel Approach to the Cross-Adaptive Auxiliary Task," Twelfth Annual Conference on Manual Control, NASA TM X-73,170, May 1976, pp. 63-71.

B-14 Sternberg, S., American Scientist, Vol. 57, 1969, pp. 421-457.

B-15 Sternberg, S., "The Discovery of Processing Stages: Extension of Donder's Method," In W. G. Koster (Ed.) Attention and Performance II, Amsterdam: North-Holland, 1969.

B-16 Sternberg, S., Quarterly Journal of Experimental Psychology, Vol. 27, 1975, pp. 1-32.

B-17 O'Donnell, R. D., "Secondary Task Assessment of Cognitive Workload in Alternative Cockpit Configurations," Higher Mental Functioning in Operational Enviroments, AGARD-CP-181, April 1976, pp.C-10-1 to C-10-5.

B-18 Merhav, S. J., and O. Ben Ya'acov, "Control Augmentation and Workload Reduction by Kinesthetic Information from the Manipulator," IEEE Transactions, SMC-6(12), December 1976, pp 825-835.

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| NASA CR-166314 | | |

| 4. Title and Subtitle | 5. Report Date |
|---|---|
| TECHNICAL APPROACHES FOR MEASUREMENT OF HUMAN ERRORS | May 1980 |
| | 6. Performing Organization Code |
| 7. Author(s) Warren F. Clement, Robert K. Heffley, Wayne F. Jewell, and Duane T. McRuer | 8. Performing Organization Report No. STI TR-1156-2 |
| | 10. Work Unit No. |
| 9. Performing Organization Name and Address<br>Systems Technology, Inc.<br>2672 Bayshore-Frontage Road, Suite 505<br>Mountain View, California 94043 | 11. Contract or Grant No. NAS2-10400 |
| | 13. Type of Report and Period Covered Contractor-Report-Final |
| 12. Sponsoring Agency Name and Address<br>National Aeronautics and Space Administration<br>Washington, D.C. 20546 | 14. Sponsoring Agency Code |

15. Supplementary Notes
Technical Monitor: David C. Nagel, Mail Stop 239-1, NASA Ames Research Center
Moffett Field, CA 94035 (415)965-6170 FTS 448-6170
See Also NASA CR-166313 and NASA CR-166315

16. Abstract

Human error is a significant contributing factor in a very high proportion of civil transport, general avaition, and rotorcraft accidents. Finding ways to reduce the number and severity of human errors would thus appear to offer promise for a significant improvement in aviation safety. Human errors in aviation tend to be treated in terms of clinical and anecdotal descriptions, however, from which remedial measures are difficult to derive. Correction of the sources of human error requires that one attempt to reconstruct underlying and contributing causes of error from the circumstantial causes cited in official investigative reports. Relevant measurements based on a comprehensive analytical theory of the cause-effect relationships governing propagation of human error are indispensable to a reconstruction of the underlying and contributing causes. This report presents the technical details of a variety of proven approaches for the measurement of human errors in the context of the national airspace system. Primary emphasis is on unobtrusive measurements suitable for cockpit operations and procedures in part- or full-mission simulation. Procedure-, system performance-, and human operator-centered measurements are discussed as they apply to the manual control, communication, supervisory, and monitoring tasks which are relevant to aviation operations.

| 17. Key Words (Suggested by Author(s)) | 18. Distribution Statement |
|---|---|
| Human behavior<br>Perception<br>Pilot error<br>Human factors | Unclassified - Unlimited<br><br>Subject Category - 05 |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price* |
|---|---|---|---|
| Unclassified | Unclassified | 139 | |

*For sale by the National Technical Information Service, Springfield, Virginia 22161

End of Document