AFFTC-TR-82-5

PROCEEDINGS OF THE WORKSHOP ON FLIGHT TESTING TO IDENTIFY PILOT WORKLOAD AND PILOT DYNAMICS

AFFTC

Edited by:

MICHAEL L. FRAZIER
Engineering Psychologist

ROBERT B. CROMBIE
Captain, USAF

MAY 1982

FINAL REPORT

THIS DOCUMENT HAS BEEN APPROVED
FOR PUBLIC RELEASE AND RESALE:
ITS DISTRIBUTION IS UNLIMITED.

AIR FORCE FLIGHT TEST CENTER
EDWARDS AIR FORCE BASE, CALIFORNIA
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE

This technical report, AFFTC-TR-82-5, was submitted under Job Order Number 99860000 by the Commander, 6520 Test Group, Edwards AFB, California 93523. It has been reviewed and is approved for publication.

The editors assume sole responsibility for the accuracy and completeness of conclusions and recommendations synthesized from the formal presentations and informal discussions of the workshop. These conclusions and recommendations reflect their summary of the broad consensus of those in attendance and like the authors' opinions in the individual papers, do not necessarily reflect the views of the Air Force.

This report has been reviewed and cleared for open publication and/or public release by the AFFTC Office of Information in accordance with AFR 190-17 and DODD 5230.9. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public including foreign nationals.

Editors:

Michael L. Frazier
MICHAEL L. FRAZIER
Engineering Psychologist

Robert B. Crombie
ROBERT B. CROMBIE
Captain, USAF
Flight Dynamics
Laboratory Liaison

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
AFFTC-TR-82-5

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
PROCEEDINGS OF THE WORKSHOP ON FLIGHT TESTING TO IDENTIFY PILOT WORKLOAD AND PILOT DYNAMICS

5. TYPE OF REPORT & PERIOD COVERED

6. PERFORMING ORG. REPORT NUMBER
AFFTC-TR-82-5

7. AUTHOR(s)
Editors: Michael L. Frazier
Captain Robert B. Crombie
(Flight Dynamics Laboratory)

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
6520 Test Group/ENAH
Air Force Flight Test Center
Edwards Air Force Base, California 93523

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
65807F

11. CONTROLLING OFFICE NAME AND ADDRESS

12. REPORT DATE
May 1982

13. NUMBER OF PAGES

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and resale; its distribution is unlimited.

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

18. SUPPLEMENTARY NOTES

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

Work Measurement Models
Man Machine Systems
Identification
Automation
Flight Testing
Simulation
Attention
Flight Maneuvers
Human Factors Engineering
Control Systems
Display Systems
Flight Deck
Pilots
Flight Instruments
Performance (Human)

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

These proceedings contain more than 40 papers or abstracts of presentations made at the workshop held at Edwards AFB, California on 19-21 January 1982. Section I contains overview papers on the subjects of pilot workload, pilot dynamics, and flight test requirements. Included in this section are papers describing the various mission environments encountered by advanced aircraft

20060223367

and design procedures used to improve pilot-vehicle performance capability. Section II addresses the quantification of pilot workload using measures of spare mental capacity, subjective ratings and opinion, and pilot physiology. Section III covers the identification of pilot dynamics and task performance including the measurement of pilot performance, the modeling of pilot dynamics, and the collection of flight test data.

General conclusions of the workshop were that workload is a multidimensional concept involving complex relationships among the pilot, the vehicle, the tasks, and the environment. Each aspect must be carefully considered in order to effectively assess pilot workload in flight. Workload should be specifically addressed as an important parameter for decision making during the evolution of design, development, test, training, and transition to operational status. Pilot dynamics and task performance complement the measurement of pilot workload. A standardized set of objective and subjective workload and task performance measures and procedures needs to be defined and introduced. Further periodic interdisciplinary conferences and flight research were recommended.

PREFACE

SPONSORS. The Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics was held at the Edwards AFB Officer's Club 19-21 January 1982. The workshop was sponsored by the Air Force Flight Test Center in conjunction with the Flight Dynamics Laboratory, NASA-Dryden, the American Institute of Aeronautics and Astronautics, and the Human Factors Society.

NEED. Advances in avionics and control systems have increased the mission capability of military and civil, fixed- and rotary-winged aircraft. The challenge to designers of advanced aircraft systems now is to make the crewmember's job easier, i.e., to allow him to precisely control the aircraft and succesfully perform the mission while maintaining adequate mental and physical reserve capacity to handle emergencies. This design problem calls for all parties involved in the research, development, test, and evaluation of advanced aircraft to understand the elements that influence the efficient interaction between the pilot and aircraft. Fundamental to this understanding are quantitative measures of pilot workload and performance, and descriptive models of the pilot's dynamics. Currently such measures and models are not mature enough for reliable use in the testing and development of advanced aircraft. A forum is needed to gather the proper expertise in the areas of flight testing, workload measurement, pilot-in-the-loop dynamics, and applications so that the state-of-the-art in these areas can be understood and approaches can be recommended that will develop flight-worthy measures of pilot workload and flight-validated models of pilot dynamics. The present workshop, the fourth in an annual series exploring specialized areas of flight testing, was structured to meet this need.

OBJECTIVE. The objective of this workshop was to bring together technical experts working with various measures of pilot workload and performance and with various models of pilot dynamics to meet face-to-face with the flight test and applications community and together to define:

1. The need for and applications of flight-worthy measures of pilot workload and flight-validated models of pilot dynamics,

2. The state-of-the-art and current problems to be overcome in order to measure pilot workload or task performance in flight or to tailor flight tests to identify pilot dynamics, and

3. Recommended ways to solve these technical problems.

PROGRAM. The program consisted of three days of presentations and informal discussions tailored to the objectives of the workshop. The first day was intended as an overview of the many facets of pilot workload and pilot dynamics. The second day featured presentations dealing with flight testing for pilot workload. The final day agenda addressed topics related to flight testing to identify pilot dynamics and task performance.

PROCEEDINGS. This volume is divided into three sections corresponding to the three days of the workshop. It is a compilation of written material submitted by individual authors for publication. A few papers not presented at the workshop are included because of their direct bearing on the workshop theme. The editors wish to thank all the authors who contributed to the quality and success of the workshop.

ACKNOWLEDGEMENT. The editors are very grateful for the assistance given by the individuals named below both in the conduct of the workshop and in the compilation of these proceedings. Indeed, without the dedicated efforts of everyone involved, the workshop could not have taken place. Special thanks go to Lt Col Paul B. O'Connor for his loyal support from the very inception of the workshop. The following individuals also contributed to the success of the workshop:

1. Edith Arnold
2. Don Berry
3. Janet Brower
4. Cy Crites
5. Lt Phillip Delaney
6. Tracie Downey
7. Ruth Gwin
8. Lt Dennis Hines
9. Jeff Holland
10. Gloria Humphries
11. Deborah Ivory
12. Gerry Jones
13. Lt David Madsen
14. Lt Col Dave Milam
15. Deborah Mummaw
16. Alfred Phillips (Technical Director AFFTC)
17. Carole Shaw
18. Suzan Stephens
19. Fred Stoliker (Former Technical Director AFFTC)
20. Flo Swartz
21. Debbie Thompson
22. Belinda Wickes

The findings of the USAF Night Attack Workload Steering Group will be of interest to the readers of this report. They will be published this year in three volumes by the Aeronautical Systems Division (ASD/ENEC), Wright-Patterson AFB, Ohio 45433 as ASD-TR-5002.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Concluded)

THIS PAGE INTENTIONALLY LEFT BLANK

SECTION I

OVERVIEW OF PILOT WORKLOAD, PILOT DYNAMICS, AND FLIGHT TEST REQUIREMENTS

A PILOT'S PERSPECTIVE ON WORKLOAD IN SINGLE SEAT FIGHTERS

DAVID W. MILAM, Lt Col, USAF
Director, F-16 Combined Test Force
Air Force Flight Test Center

INTRODUCTION

The recent loss of four fine young pilots at a Thunderbird practice in Nevada makes our job of analyzing pilot workload even more critical. It is obvious to me that we have a lot to learn about the manner in which the brain functions when the pilot is doing high gain tasks. We should set a goal for ourselves of learning enough about mental and physical workload to make the probability very remote of ever having a similar accident.

Pilot workload in single seat fighters is obviously increasing. That pilot workload includes physical workload and mental workload. The physical workload has steadily increased. In the past, the F-4 could pull 7.33 g's. Today the F-16 can sustain 9 g's, and in the future the AFTI/F-16 will add 2 lateral g's to the aircraft maneuvering capability. The impact of that physical workload on mental workload and task accomplishment needs to be determined. Mental workload demands have increased by a gigantic amount due to the increase in the amount of information displayed to the pilot and the increased complexity of modern systems. The F-4 has a radar & lead computing sight. The F-16 has a heads up display, stores control panel, fire control computer, and sophisticated radar. We are also forced to fly faster and lower to evade sophisticated ground based threats. The new MSIP III F-16s will have two multi-function displays, wide field of view hud, internal ECM, global positioning system, navigation pod, and an increased weapons capacity of radar missiles and 30 mm gun. The next generation fighter will have to be designed to kill tanks at night in European weather and battle for air superiority against overwhelming odds. Pilot workload must be one of the major factors to consider as new weapons systems are developed.

The decision process for designing and building new systems is complex and good decisions are based on a good analysis of the tradeoffs between cost and benefit. In order to determine the relative balance of gain and cost, there are several impacts which need to be known:

Cost
Performance
Maintainability
Reliability
Pilot Workload

If a system cost nothing, improves performance, decreases workload, and does not change maintainability or reliability then the decision to incorporate is easy. That combination never occurs, however, and we need to accept the responsibility to provide a complete picture of cost and benefit tradeoffs to decision maker.

Cost can be determined with at least some degree of confidence by signing a contract. Hidden costs always exist but within reasonable limits, costs can be determined.

Performance can be determined also. During flight test or simulation we can assess the total response of the pilot and the vehicle by measuring bomb scores or evaluating landing accuracy.

Maintainability is not easily measured and we do not always structure tests to evaluate maintainability. There are, however, some traditional methods of measuring the total impact by measuring the man-hours required to keep a system flying.

Reliability is a top priority. The long term impact on lift cycle cost and maintenance man-hours is significant. Additionally, the decrease in compat capability of aircraft having sophisticated systems is large. The combat effectiveness of an F-15 without an operable radar is very low. It is possible, however, to measure mean time between failures (MTBF) and have a general picture of reliability.

Cost, performance, maintainability, and reliability have some definite standards of measurement which allow comparision for good decision making, however, pilot workload does not presently have that definitive measurement. It is clear that future decisions need to be based on an accurate, complete workload measurement. The methods of measurement which seem to be possible are quantitative measurements and pilot qualitative opinion.

CHALLENGES TO WORKLOAD DETERMINATION

There are several considerations which I believe impact pilot performance and workload. I have included in the Appendix a discussion of flying as a skill task where training, hand-eye coordinating, conscious versus subconscious information processing, and individual techniques are involved. These concepts form the foundation upon which I will base the following challenges, as I see them, to workload determination.

Research to measure mental workload as well as physical workload is very important because pilot qualitative opinion is not always consistent. Remember, pilots are unique. Just as we measure the short period damping of an airplane, it would be nice to measure physical and mental workload with a laboratory proven, measureable parameter. It would be nice to have some unobtrusive physiological measurement to quantify pilot workload. Two different measurements (such as heart rate and eye movement) will probably be needed to distinguish between physical and mental workload. This distinction is needed by decision-makers in the design of new systems. For example, the separate effects of side-acceleration restraints and various display formats will need to be evaluated during AFTI/F-16 bombing tests.

In reality, however, the measurement of workload is still accomplished primarily by pilot qualitative opinion. A pilot estimates his workload. For that workload estimation to be meaningful, it is mandatory to take the average of several pilots. For the estimation of workload to be consistent, then, we need to strictly control the tests and use the same rating scale.

Mr Tom Twisdale has written many documents on how to accomplish flying qualities evaluations. He structures tests in a precise manner, measures performance, and collects pilot ratings. His primary (but not his only) rating scale is the Cooper-Harper scale which combines performance and workload. After our recent flight tests to evaluate avionics subsystems there is a question in my mind about the utility of measuring performance and workload together on the Cooper-Harper rating scale. Most test pilots do not use the C-H scale properly but tend to give a relative numerical rating from 1 to 10 where perfect is 1 and uncontrollable is 10. A modified subjective rating scale method should be developed to allow the pilot to evaluate workload and performance separately. Such a scale could be validated during flying qualities evaluations along side the Cooper-Harper scale. The scale would then be of great value to the evaluation of avionics, flight control, and other subsystem changes and their effects on total system performance.

Workload measurements, whether subjective or objective, should be available much earlier in the systems design process so that design options can be intelligently considered. If cost, performance, maintainability, reliability, and workload impacts are known very early on all design options, then good decisions may be made by rational, motivated leaders. However, the "real world" is that design decisions are often made before hardware is yet available for pilot evaluation. Most pilot evaluation of new systems is currently done in the flight test phase where the number of design options left for evaluation is limited. We would gain a great deal by accomplishing meaningful evaluations at an early stage in the design process. At the risk of alienating all of the pilots who may read this, that early evaluation means that piloted simulations need to occur early in a systems development program. An early estimate of the impact of various design options will be the result. Additionally, many more options can be considered because simulation is much less expensive and because there is more simulator time available than flight time. The present radar and fire control computer evaluations in the F-16 are outstanding examples of flight tests which need to be supplemented by good piloted simultation efforts. Ground simultation of those systems would allow changes to be evaluated much more efficiently and allow the design process to converge on the best solution more rapidly.

It is mandatory that several pilots be involved in the piloted evaluation and development of new systems whether that is accomplished in a simulator or on an airplane. During the control law tests which I have managed on the AFTI/F-16 and on the basic F-16, the final configuration as developed by several pilots was never the same configuration as would have been developed by a single pilot. In each case, however, the final configuration was always liked by all of the pilots. During all of those evaluations, every pilot had a VETO and every pilot had an equal vote in the development process. In most cases, unbiased pilot input was achieved by requiring the pilot to write a written report before talking to other evaluation pilots. An optimum number of pilots apperas to be between four and eight. Valid piloted simulation evaluations, then, require several pilots participating together in a strictly structured design evaluation process. Of course the simulation inself and the piloting tasks involved must also be realistic.

There is another challenge that we at this workshop need to accept. Pilot-vehicle interfaces are no longer limited by hardware. The limit can be the pilot (workload saturation) if we don't design good systems. Those systems need to communicate with the pilot in new and innovate ways: voice, color, etc. We need to incorporate state estimators to determine when we display what piece of data. We need to accept as fact the uniqueness of pilots which demands that we include flexibility in the design to accommodate individual differences.

The flight test community is limited in many ways so it is critical that we provide good pilot-vehicle analysis early in the design process. We are:

(1) Time limited - only so many hours in the day.

(2) People limited - a finite number of engineers and psycholinguists limits us in our ability to look at every possible new idea.

(3) Dollar limited - we cannot afford to flight test all of the different design options available to us.

Because of these real limitations we must:

(1) Study the pilot/aircraft system at an early point in the design process through meaningful/realistic tests.

(2) Prioritize the data displayed to the pilot because of the wide range of data available for display.

(3) Provide meaningful workload and performance data for decision makers early in the design process.

(4) Develop systems which accommodate individual pilot differences.

(5) Develop systems which give good total system performance.

APPENDIX

SKILL TASK ACCOMPLISHMENT RELATED TO PILOT TECHNIQUE AND AIRSHOW SAFETY

Human skill task performance needs to be understood by all involved in workload measurement.

Skill tasks are learned by the human through a process which seems to be consistent whether the skill task is flying an airplane or hitting a baseball. Several aspects of this process are:

(1) The mind works on two levels: conscious & subconscious:

The conscious mind does only one task at a time while the subconscious mind accomplishes many tasks simultaneously. The conscious mind time shares. Every event which happens to you is stored in the subconscious. Skill tasks are accomplished best when the task is done at the subconscious level.

(2) Tie between Foveal vision and conscious:

The strong tie between Foveal vision and the conscious mind is important because it establishes the necessity to have a good visual cue in each skill task: "watch the ball," "watch the rim," "stare at the target." Subconscious hand-eye coordination is developed in this way.

(3) Proper visual cue and practice is the key to skill task performance:

The key to skill task performance is to use the proper visual cue(s) and then to practice a great deal to get a good data base in the subconscious mind ("program core"). Of all the cues that we use to accomplish a skill task, the visual cue is by far the most important. Probably 90 pct of all cues for doing skill tasks are visual cues.

(4) Each pilot has a unique style:

It is extremely important to understand that each human pilot is unique. Every pilot flies differently. Traditionally, we have tended to imagine the standard pilot or the great pilot. He doesn't exist. "The world's greatest fighter pilot" is found only at the Nellis AFB Officers Club Bar.

Pilot performance of skill tasks in flight very closely parallels that of various other skill tasks: flying, tennis, basketball, skiing, riding a bicycle, etc. Most of the mental work in these tasks is done on a subconscious level. The subconscious work is done in one part of the brain and the conscious work is done in another part. The FOVEAL vision of the eye (the central vision or sharply focused vision) is processed by the conscious part of the brain. The subconscious part of the brain processes both the FOVEAL and the peripheral vision. The conscious part of the brain performs the reasoning or analytical functions and it only does one thing at a time. In essence it "time shares" various tasks and problems and takes some finite amount of time to accomplish each separate calculation, decision, or analysis. The subconscious mind, on the other hand, responds to information and gives out a large number of commands simultaneously; it can do many different functions at the same time. With these things in mind, it is obvious that the key to all skill tasks is to do most of the work on the subconscious level. In each separate skill task the person must use the correct visual reference and then accomplish a large amount of practice to establish a large data base in the subconscious mind. I like to refer to that practice process as "programming core" because of the obvious similarity with digital computers. In each separate skill task there is a significance to the "appropriate sighting reference." Firstly, the conscious mind becomes occupied when the individual concentrates on that sighting reference and that allows the subconscious mind to do most of the task. Secondly, in every skill task the mind needs some visual reference as a foundation from which to make adjustments. In flying that visual reference is pitch attitude. In tennis the great (even good) player watches the ball very intently. In basketball the great player with the consistent jump shot always focuses instinctively on some specific part of the basket. The great hitters in baseball watch the ball so intently that they can actually see the spin of the ball. It becomes imperative in all of these tasks that the individual leave the majority of the mental work on the subconscious level; when we try to elevate individual parts of the task to the conscious level then we slow down in our performance. An outstanding player in any sport is the individual who has supreme confidence in himself and in his coach so that he leaves most of the task on the subconscious level. If the coach tells him to do something in a specific way then he practices that without question until he has "programmed" the specific parts of the task on the subconscious level. When this individual is asked how he does a task, he then tends to answer by saying that he does what comes naturally. There are some good players who try to dig into all the separate parts of a task, and by doing this, they elevate each part of the task to the conscious level. They can attain some high performance level but they never quite become the greatest. They are just a little bit slower than the "natural athlete" which was described first. This second individual becomes the really great coach even though he was never the best of his field as a player. This fact is logical because the natural athlete never did study the intricacies of a task while the good coach has studied each task in exhaustive, thorough, comprehensive detail and can therefore teach others. By elevating the task to the conscious level he understood it better, but he also prevented himself from being absolutely great. When a coach who has studied tasks in detail teaches his players, he speaks from a sound foundation in fundamentals and urges them to master the proper skills. As his advice leads them to success, the players gain faith in themselves and their coach. They leave most of the skill task in the subconscious level as they should to gain great success.

You may now ask, "How does this impact flying airplanes?" Just as in all sports, there are parts of the flying task which require skill while some things involve brute force. We need to use all of the cues available and we need to have a good visual cue. When we are landing the airplane or accomplishing air/air tracking, then we need to practice those tasks several times to "program core." The conscous mind "time shares" its tasks/ decisions; therefore, it becomes important to establish some priority of things that need to be consciously monitored or checked. Hence the standard instrument cross check becomes very important to flying good instruments. The conscious mind/Foveal vision spend most of their time on the primary visual cue (ADI). Other parameters are reviewed on a time-sharing basis. The check becomes:

$$\begin{tikzcd} \text{A/S} \arrow[r, leftrightarrow] & \text{ADI} \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] & \text{ALT} \\ & \text{HSI} & \end{tikzcd}$$

In addition, the engine parameters, VVI, etc, are checked less frequently. The key to A/A gunnery is also obvious. Focus your eyes and conscious efforts on the target primarily. Let your subconscious mind do most of the work. "Time-share" tasks in the correct sequence: Radar lock-on, get in the target's turning plane, start the pipper drifting toward the target, accomplish fine tracking, shoot at the right time.

In all of these tasks it is very important that the mind and the body be in tune. For all of these skill tasks it is extremely important that we be relaxed but alert. Remember, have your inner self and your outer self in tune: Both your physical self and your mental self must be relaxed but alert.

There are times when all of these facts will help you to fly an airplane better but it will not cause your death if you forget them. However, we periodically place ourselves in situations where the smallest mistake can kill us: Air/ground gunnery, formation, or airshows. In an airshow, we are flying extremely close to the ground and a mistake might kill us; it becomes very important to determine what parameters will kill you and what parameters will simply make the show look a little less impressive. For each maneuver we establish a priority list of parameters that need to be checked. Most of the flying task has been given to the subconscious mind while the conscious mind focuses on the correct sighting reference (aircraft attitude primarily) and time-shares checking other parameters based on the priority list established before flight. If we "program core" by practicing the airshow several times, the airshows can be made very safe.

There can be many distractions which disturb the established routine. Anything that is external to the normal thought process may get elevated to the conscious level because "core" doesn't know what to do about this "unprogrammed" event. Remember that the conscious mind only does one thing at a time (it "time shares" separate tasks) and that the normal parameter check sequence can be broken by an unprogrammed event. If one of those parameters that you miss because of the disturbance is a critical parameter, then that disturbance might kill you.

In addition to external distractions, there are internal distractions. If your mind is thinking about something other than the airshow then you might miss a critical parameter. It is extremely important to cleanse the mind of all distractions. Most pilots will do this by going into isolation for a period of time prior to their airshow. By mentally reviewing each maneuver and mentally flying the airshow, then the pilot can cleanse his mind of other events which might cause a fatal distraction. All airshow pilots should instruct their support crew and friends about the effect of a mental distraction; they should not bring problems to the pilot when he is in his airshow preparation. The pilot should know the possible fatal impact of mental distractions and should cleanse his mind of any last minute problems before he goes to fly the airshow. Any airshow pilot is "tickling the bear" if he stays in the hospitality suite until just before his flight. Such overconfidence leads to filling the conscious mind with unneeded thoughts to the conclusion of safety-critical cross checks. I believe many airshow fatalities could have been avoided by following this checklist:

(1) Prepare your airshow and establish a parameter priority list for each maneuver.

(2) Practice until "core" has been adequately "programmed" to give a safe demonstration.

(3) Go into isolation before the airshow to cleanse the mind of extraneous distractions.

(4) Get your inner self and your outer self (mental and physical) in tune. You must be relaxed but alert.

(5) Do not allow external or internal distraction to disturb your conscious mind in its established parameter check sequence.

ARMY WORKLOAD RESEARCH AND DEVELOPMENT REQUIREMENTS

Richard S. Dunn, PhD

US Army Research & Technology Laboratories
Ames Research Center
Moffett Field, California 94035

This article presents a discussion of the principle current Army requirement for research and methodology development in pilot and aircrew workload measurement. Emphasis is given to the shifting impact of operator task demands and to the utility value of workload methodology in guiding design and development activities. The discussion focuses on requirements which are especially acute in the system acquisition process involving development of new helicopter crew stations. Conclusions and opinions offered by the writer suggest both a conceptual orientation and a general approach to meeting the requirements.

Improved workload measurement methodology is needed by the Army for several purposes in the helicopter system acquisition process. Although there may be other potentially fruitful applications for the workload concept, the focus here is on system acquisition. In many ways the Army is in a transition period with respect to helicopter development and application. Both missions and crew task demands are undergoing rapid and extensive change. Relatively recent changes include adoption of the anti-armor role as the primary mission for the attack and scout helicopter team, and the expansion of night and all weather capability to meet the needs of continuous operations. Extended development of terrain flight tactics and long-standoff target acquisition capability are also sources of new task components. In addition, mission requirements for air-to-air combat with helicopters now appear to be gaining importance.

Tactical helicopter system designers are incorporating many advanced subsystems for crew job aids or for automation of previously manual tasks. The trend is already highly visible: disengagement of the pilot and crew from routine aircraft operation through use of more reliable or automatic basic subsystems (powerplant, transmission, electrical supply etc.) and increasing task engagement with mission related subsystems (target acquisition, weapons etc.). Mission equipment packages are increasing in complexity and operator attention demands. Many new subsystem elements are recent developments with which the Army developers and users have little prior experience. This means that dependence on convention or prior practice is impossible, and in many cases the task demands imposed on the operator in system operation are unpredictable. Workload consequences are uncertain during development and may not be fully known even after completion of the developmental test cycle. At important system development choice points, equipment tradeoffs are made with reliable cost, weight, space and development risk data and only hesitant workload and mission effectiveness forcasts or guesses. System project managers cannot be blamed for often selecting the design approach with lower cost, weight, size and risk even though the end result may be untenable demands for high performance on the part of system operators.

Since poor system design or excessive mission demands all result in high operator workload there has been increasing general attention to the subject. Ironically we now find ourselves ready to conclude that the workload idea has little merit in the scientific process of theory and model building attempting to account for human behavior. Because of overlapping meaning with numerous well established behavioral processes or variables, an operational definition containing much of the everyday sense of the term has not been forthcoming. But it is suggested here that the popular appeal of the term and renewed attention to good human engineering in military systems need not be wasted. Aside from its potential explanitory and scientific value, which may be in doubt, the workload idea may yet have a high degree of utility value when developed for application in the system acquisition process.

To reach this potential a standardized measurement methodology must be developed and refined. One main point of this discussion is that the methodology research should be related in a direct way to the purpose intended for the measurement methods. An explicit statement of the need and use for measurement methods is a key requirement in selecting an approach to their development. The conceptual orientation suggested here is that workload measurement methods should be viewed as a product, the result of a development effort in applied psychology. The product in this case is for use by developers seeking more effective methods for helicopter weapon system design, development and evaluation.

A product oriented effort to tailor systematic workload measurement methods to a specific end purpose really amounts to a psychological test development task. It should be noted that there are both benefits and drawbacks to this conceptual orientation. The benefits are that developers can start with a careful evaluation of the requirements and work backwards (at least in the first stages of test development). This is not the normal approach in research psychology, being a strategy usually available only to engineering development. A second benefit is that the explicit techniques which will become the standardized measurement methodology probably already exist in their scientifically oriented format along with a body of useful data. The biggest drawback to the suggested orientation is the need to recognize the limited nature of the objective. Scientific merit and universal generality will (probably) not be the result. The eventual methods developed may fall short of the ideal for measurement tools and will require careful application, judicious interpretation and some active protection from misuse. Also, the numerous other fields in which the workload idea may be useful (personnel selection, training, system certification etc.) may not find the tailored product applicable to their problems.

Fully developed workload measures would aid the system acquisition process in many ways. To identify the practical requirements which provide constraints or determining characteristics in the test construction task a discussion of the ultimate uses and benefits to be obtained is in order. Workload measurement could provide a procedural basis and framework to organize human engineering efforts. It would permit empirical comparisons between systems to develop normative data reflecting mission contingency reserve requirements and to establish sustainable workload levels for normal mission operations. The same information would aid in showing the relationship between the quality of system design or operation and mission effectiveness. The methodology would permit systematic consideration of workload in all design development and evaluation phases. By providing a basis for specfication and statement of work entries, it would permit workload for specific system functions to become an engineering parameter in system design tasks. At least for some system functions workload would have the status of a system characteristic and could be weighted fairly in tradeoffs with other parameters such as weight, speed, cost, etc.

These end objectives imply a generality and standardization level that would permit the workload measures to be applied to any combination of system or subsystem crewstation, operator, and mission component or task. To be most useful in design, development and evaluation steps the measurement procedures should characterize crew task demands which are a function of system design quality and which are under the control of designers or developers. In application, workload measures would apply to specific tasks or functions; the assessment would not be a universal evaluation of a system for all of its potential uses. The workload measures need not apply to workload determinants which are outside the control of system designers (except to the degree necessary to establish experimental control when required). Another implication is that individual operators are not the subject of evaluation. So long as the operator's skills, motivation, stress levels, risk taking behavior, operating strategy, fatigue state and etc. are within the range of the intended user population, these factors and the individual differences so strongly present within them are not the subject of the measurement procedure.

This overall requirement, joint evaluation of equipment and tasks but not individual operators, frames the basic test development task. In addition, the results should permit users to focus on various workload determinants associated with the equipment or its operation in flexible ways depending on the needs of specific applications. Workload measurement procedures with more than one measurement or result offer the possibility of a diagnostic type of use in which the task components or equipment components associated with workload issues could be evaluated. By relating the measurement procedures discussed here to more rigorously anchored behavioral data from laboratory research and from subsystem development data, some capability to predict workload for combinations of both equipment (more than one subsystem) and tasks (various system functions) serially or together may emerge. Using this capability the system designer could control the workload consequences of the overall system design in order to avoid peak overloads, minimize errors due to excess workload levels, control operator skill, training and proficiency requirements, establish balanced criteria for equipment and mission reliability and predict the suitability of failure mode back- up systems.

Up to this point the discussion has identified a test construction task with limited and specific objectives. The resulting procedures will not reveal universal truth about workload for use in theory or model building efforts. They will be a set of operational definitions to establish the procedure for and explicit meaning of workload measurement. The result may not be useful procedures outside the realm of system acquisition. What can be hoped for is that through construct validation the resulting measures may eventually attain a cohesive sensible meaning in their own way that is consistent with other concepts in engineering psychology. Whether or not this occurs the measures developed may be formalized and standardized. They may form the basis of an extended design-development-test technique that augments analytical techniques such as task analysis and is practical for inclusion in system development work statements, specifications and standards. For any given procedure, initial success may be limited until usage provides sufficient experience against which to judge individual outcomes. Construct validation requires converging lines of evidence from different applications. At the outset, findings may only generate rule-of-thumb guidance and surely will not provide all the potential benefits discussed above until fully developed to a high level of maturity.

On the way to this perhaps utopian objective, developers must encounter the standard list of desirable properties for all psychological tests. The test procedures and its outcomes should be reliable, efficient, continous, sensitive, and non-intrusive. Only the last of these will be discussed. In the present context a non-intrusive measure would not alter the process in a way that would spoil the measurement. In workload research this constraint is usually taken to mean several things that it actually may not mean. Strong changes in operator strategy or response style as a result of workload measurement introduction surely must be avoided. But it may not be possible, or desirable, to introduce workload measurement without any change in the operators task demands, without the operators awareness, or without some degree of conflict or overlap between the tasks underway originally and those associated with the workload measure. This is particularly true if an ordinary secondary task approach is employed. In any event, all that is really required to satisfy the non-intrusive measure constraint is reasonable assurance that the process being measured is not so altered by the measurement, itself, that the results are invalid for the purposes intended.

There are several desirable features the writer would add to the standard list. First, the methodology should have high face validity both to test subject system operators and to the acquisition process managers. These are user acceptance factors that are essential for cooperation and application of findings. Also, specific techniques should be easy for both evaluators and test subjects to learn. For practical reasons automatic administration and scoring would also be desirable characteristics for any specific set of measurement procedures.

The next consideration in reviewing this test construction task is to survey the known pitfalls awaiting unwary developers. For workload measurement there are discouragingly many. The most important issue has already been discussed and involves lack of specific orientation to the application usage. Scientific efforts applying a reductionistic-analytic-experimental approach usually result in a switch in topics from workload to learning, vigilance, motivation, fatigue, stress, individual differences, risk taking, and so on. The level of reductionism attempted in a composite measure to characterize workload must be appropriate to both the setting of the application and the usage for the results. Various practical problems in system testing provide serious constraints which limit test data observation opportunities, and which may introduce some of the problems associated with field testing. Low experimental control, unreproducible scenarios, varying external factors which introduce situational conditions in system operation, and, at least at the outset, the absence of firm validation criteria must all be considered by the workload methods developer. Other problems or pitfalls are associated with the optional timing of many system operation tasks, the need to separate workload from multiple system outputs - some of which are antagonistic or strategy dependent, and the potential need to deal with a very large set of systems functions which may include back-up or failure mode operations. For many conditions of widely varying workload the only difference in system output may be hazard or risk effects which are unquantified. In the end, the relationship between workload and system effectiveness will have to be made clear by empirical demonstration or the uses planned for the workload information will not emerge.

For sure, there are plenty of problems to overcome in the application of workload measurement to aid system development. But the need is strongly present and well recognized. The potential benefits are extensive and important. Our behavioral technology surely has the necessary tools to attack the problem in a productive way. This convergence of motive, means and meaningful consequence for success is all that is really needed to justify a serious development effort. With the product oriented approach suggested here a set of procedures tailored to the end usage can be developed that will provide many, perhaps most or even all of the benefits suggested. Workload methodology research can make the strongest contribution to Army aviation system development capability by meeting the challenges of this task.

WORKLOAD REQUIREMENTS OF THE HELICOPTER ANTISUBMARINE WARFARE MISSION

by

LCDR M. A. COSGROVE, USN

The pilot workload in the helicopter Antisubmarine Warfare (ASW) environment is unique in many aspects, particularly when coupled to the harsh environment of at-sea night operations. For the purposes of this discussion, I will primarily address the tasks involved in piloting the Navy's SH-3H helicopter.

The SH-3H "Sea King" helicopter is equipped with a mechanical flight control system augmented by a hydraulic boost system and Automatic Stabilization Equipment (ASE). Additionally, it has an "automatic coupled approach" feature and a barometric attitude hold system. It is similar to the Air Force's CH-3E helicopter, and it is manned by a crew of four: a pilot, copilot, and two sensor operators. Side-by-side seating is provided for the pilots in the cockpit, while the sensor operators man their equipment in the cabin section of the aircraft.

The SH-3H has a maximum airspeed of 120 KIAS with an average mission endurance of 4.5 hours. It is basically limited by a restricted power available (critical in high temperature/density altitude regimes), and it lacks radar, ECM, and inertial navigation systems. The large size of the cabin which allows pilots to walk around during long missions, its boat hull which will allow water landings in certain situations, and a new tactical navigation system (ASN-123) contribute to the flexibility of the SH-3 helicopter.

In order to accomplish its primary ASW mission, the SH-3H makes use of a variety of sensors including a dipping sonar, Magnetic Anomaly Detection (MAD) equipment, and various types of active and passive sonobuoys. It is operated primarily from the decks of the Navy's aircraft carriers in day/night all-weather conditions; and it is frequently flown off the decks of smaller aviation-capable ships such as the USS SPRUANCE class destroyer (DD-963). The SH-3 is often employed under Emission-Controlled (EMCON) conditions without any navigation aids (i.e., TACNAV) at ranges in excess of 100 NM from the carrier. Although it is primarily used in the localization and attack role, it is capable of performing limited open-ocean searches. The SH-3H is also tasked with performing surveillance, search and rescue, and utility missions.

The pilot workload in the cockpit rapidly builds up during the following three major phases of the mission:

a. Tactical scenarios which involve the use of multiple sensors, coordination with other aircraft, and multiple frequency communication links.

b. Night sonar dipping maneuvers.

c. Night shipboard approaches and landings.

During tactical scenarios, the pilot is required to devote a major portion of his concentration in tracking an enemy submarine, while relying on his past training and motor skills to maneuver the aircraft to accomplish this objective. The pilot must prioritize his subtasks and time-share his focus of attention while utilizing up to three radios and internal communications; updating his position and the positions of the submarine, his wingman, or supporting assets; coordinating the use of the navigation system; and frequently referencing tactical publications. All these subtasks must be accomplished while safely flying the aircraft throughout its flight envelope.

The night sonar dipping operations pose unique problems to the helicopter pilot, which are not present in fixed-wing operations. The pilot must fly a profile which transitions the aircraft from a forward flight level attitude to a zero ground speed, 40-foot hover condition, which places the aircraft in a nose up, left wing down attitude. Although the "automatic coupler" is designed to fly this profile with no assistance from the pilot, in actuality, tactical situations, weak Doppler return and weak ASE systems often require the pilot to manually complete this maneuver. The difficulty of this maneuver is further compounded by the lack of a low airspeed system, the lack of outside visual cues during IMC operations, and the effects of varying sea state and wind velocities. Departing the hover again requires a great deal of concentration by the pilot to successfully transition from the "hover" attitude to a safe airspeed and altitude combination in order to arrive at the next calculated position of the submarine. Depending on the submarine's speed and location, the time of the evolution may range from 6 to 20 minutes.

The final major phase of the helicopter ASW mission which rapidly builds up the pilot workload is the night shipboard approach and landing. As in the case of the sonar dipping operation, the pilot must transition the aircraft from a level forward flight condition to a precise hover over a moving deck. During the approach phase, he must compensate for weak visual cues which vary from ship to ship due to a lack of lighting standardization of the various classes of aviation-capable ships. The difficulty of the approach is further compounded by the lack of closure rate information available to the pilot. Unless the pilot is flying a Carrier-Controlled Approach (CCA), which is not available on most of the aviation-capable ships, he must rely on his copilot's vision to detect the closure rate to the ship under the restricted visual conditions mentioned previously. (NIGHT HELICOPTER APPROACH FILM SHOWN) When pilot has finally completed his approach and begins the landing phase, he is faced by many additional harsh external factors which control the difficulty of the touchdown. Ship motion, varying winds, turbulence from ship's superstructure, stack gases, and a confined landing area are factors which demand a high skill level and compensation ability on the part of the pilot. The relative motion cues are also diminished at night, even though the pilot is hovering directly over the deck. (FILM SHOWING SH-60B RETRIEVAL, ASSISTANCE TRAVERSE (RAST) SYSTEM)

Throughout all of these major workload phases of the helicopter ASW mission, several other environmental factors combine to increase the fatigue level of the pilot. High ambient noise levels accompanied by varying vibration levels (which become severe during the translational lift portions of landings and takeoffs), high cockpit temperatures in certain climates, and the multiple flight and equipment instruments which require monitoring have often induced pilot disorientation with sometimes lethal results on more than one occasion.

Present technology offers many possible solutions in reducing the pilot workload in this environment. The use of a HUD or an integrated CRT should be evaluated to reduce the pilot's instrument scan requirements and disorientation by providing command directors and an artificial horizon. Closure rate equipment coupled to a low airspeed system may be the solution for problems encountered in the night approach to landing and sonar dipping phases of the ASW mission. The installation of inertial navigation equipment will reduce the inaccuracies of present systems and should limit the accompanying pilot anxiety and frustration levels associated with solely Doppler–based navigation systems. An increased use of helicopter landing systems, such as the RAST system, should not only make the shipboard landing task much easier but will allow the helicopter to be operated in higher sea states and weather conditions.

The tasks of the helicopter ASW mission have been defined, but precise workload measurement of these tasks is still vague and, in some cases, unknown. The operational success of the ASW helicopter will depend upon the correct identification of those factors which make the pilot's task difficult, accurate measurement of pilot workload and compensation, and finally, evaluation and incorporation of mission and aircraft equipment which will improve the performance of the helicopter in the ASW mission.

01/13/82

CIVIL AVIATION AIRCREW PERFORMANCE ENHANCEMENT AND ERROR REDUCTION

PETER V. HWOSCHINSKY
ROGER NEELAND
JOHN PARK

PRESENTED TO
AFFTC/NASA DRYDEN/AIAA WORKSHOP
ON
FLIGHT TESTING TO IDENTIFY PILOT WORKLOAD
AND PILOT DYNAMICS

JANUARY 19-21, 1982

EDWARDS AIR FORCE BASE

INTRODUCTION

We have never been at a more critical stage in the need to be able to measure workload and performance in both the cockpit and air traffic control. In order to illustrate fully why this is so, I would like to describe briefly the Federal Aviation Administration's (FAA) engineering and development programs in Human Factors.

The human factors area has been the subject of much study over many years and the results of these studies have had a major impact on the aircraft and air traffic control systems in use today. In conducting our current efforts we are not starting from "scratch," but building on and improving the already high performance of the current system. The focus of the current FAA efforts is not on "knobology" or the location of displays and controls best suited to the physiology of the human being, although this is certainly an important area. Rather, it rests on areas such as the following:

1. The causes and types of human error and the impact of these errors on the safety, performance and productivity of aircraft and air traffic control system operations;

2. The definition of automation approaches that assume the continued existence of human as well as machine error and strive to reduce both the occurrence and the consequences of such error;

3. Assessment of the proper distribution of air traffic and aircraft control monitoring functions between automation systems, the controller and the pilot;

4. Determination of the appropriate interfaces between the man and the machine at each step up the ladder leading to higher levels of automation; and

5. Determination of adequate automated, semi-automated and manual system backup capabilities to permit safe continuation of system operations under a variety of conditions of human and machine system failures.

These areas of R&D are all directed toward the need to maintain and enhance the safety of the aviation system, to achieve improved performance of the system for the participants and the flying public, and to make the system more productive while constraining the cost of the system to the Nation.

Before proceeding with an overview of the FAA's human factors program, I would like to give you a short description of what the Air Traffic Control (ATC) system of the year 2000 might look like including both ground automation and airborne system capabilities.

SCENARIO FOR THE FUTURE SYSTEM

The air traffic control system in the year 2000 will continue to be a civil/military integrated system and will continue to grow in an evolutionary manner. In the future, in some airspace, services will be provided by an essentially fully automatic ground-based system, coupled with airborne automated systems. On the other hand, there will still be airspace where no air traffic control ground-based services will be provided, and none needed. We see the possibility of placing more responsibility into the cockpit, thus changing the balance of control responsibility between pilots and controllers.

Where there is full participation in ATC services and in major terminal areas, the highest level of automation will be employed to support the separation service. The basic control system will be designed to offer very high reliability with reversion capability to a safe backup automatic control system. Reliability of the automated system will be achieved by top-down redundancy at individual control facilities, and facility-to-facility backup using a flexible data communications network. The automated system will be monitored automatically also by an outer-loop backup advisory system which will independently provide information directly to the cockpit and to controllers. Another level of protection, which may also serve to build confidence in automation, may be achieved by use of cockpit traffic displays utilizing the data link information from the ground and/or an independent collision avoidance system. Flexible use of area navigation and an automated ground-based system capable of handling large numbers of variables will allow a high degree of routing freedom, both laterally and vertically. Our work on the Automatic En Route ATC system, called AERA, will serve as the basis for this system. We believe that even in this airspace, some degree of self-separation may be possible if traffic information can be provided to the cockpit in such a way that aircraft are aware moment-to-moment of the position and perhaps the intent of other aircraft, and that procedures can be worked out to assure that pilot and controller actions are fully understood and coordinated.

In the second level of service, provided in low- and medium-density traffic areas with mixed IFR/VFR operations, the service will be similar to the present system, but with automated tools such as metering, sequencing and spacing, conflict alert, and conflict resolution advisories assisting the controller.

In the third level of service, IFR operations under Instrument Meteorological Conditions (IMC) will be permitted on the basis of procedural rules in a manner similar to that currently employed under non-radar procedures. This service will be available to provide some degree of control to protect aircraft in areas where there is a limited number of aircraft on IFR flight plans.

We have defined a series of capabilities for the ATC system at the turn of the century which will, in turn, require the development and implementation of services and tools beyond those available today. Let me mention a few of those services.

The system today provides a series of services including flight plan and radar data processing, semi-automatic hand-off between sectors, automatic inter-computer communications, conflict alert, terminal minimum safe altitude warning, central flow control, fuel advisory departure service, a central military flight reservation service, Flight Service Station services to the general aviation community, and a whole range of airport ATC services.

For the future we see a system which will have the capability of essentially all-weather operations, supported by Microwave Landing Systems (MLS); en route minimum safe altitude warning; a dramatically improved weather service for the en route system, major terminals, and low activity terminals; automated Flight Service Station services; conflict resolution advisories; and Automatic Traffic Advisory Services. The automation system will be extensive and will include an automated en route metering system and conflict probe, a national integrated flow management service for en route, terminal and airport operations, conflict-free automatic fuel optimal direct routings, and fully automatic hand-off.

As additional services and tools, we expect to see an automatic air/ground communications service which will serve a variety of functions. We expect to have a more nearly automatic airport surface surveillance and control capability at major terminals, an automatic digital data weather distribution system which will draw on a new National weather network, and the capability to achieve reduced vertical separation above FL 290.

A Threat Alert and Collision Avoidance System (TCAS) will be available, based on the existing and improved Secondary Surveillance Radar System (called Mode S by the International Civil Aviation Organization, ICAO). It will render valuable service especially in over ocean areas and in areas where ATC services are limited.

We expect to see progress in permitting general aviation aircraft to operate predominantly outside the ATC system, if they choose to do so, while retaining the protection of the system where collision risk is significant.

Our work continues in establishing the optimum role for cockpit displays of traffic information, and the problems of wake vortex alleviation remain a major challenge.

The cockpit of the future will be different from today's, and we will arrive there by an evolutionary process. A number of changes are already visible in the laboratories of industry, NASA and FAA and, of course, the military. The key word is undoubtedly integration--integration of displays, integration of functions, and integration of warnings and alerts. Highly flexible cathode ray tube or plasma displays are already appearing and, when proven fully reliable, will undoubtedly be highly valuable and flexible tools for the information exchange between the pilot, his aircraft and the ground system. We will see these new display systems integrating not only the information currently available in cockpits, but also new information from a variety of sources. Among these are the performance management computers, already demonstrating the capability of providing major fuel savings. They will, in most instances, be integrated with 3D or 4D area navigation capability. New information will also appear. There will be better displays of wind shear information, data link terminals which will connect the cockpit to the company and to the air traffic control automation system via digital communications.

Threat alerts on nearby aircraft and collision avoidance information will be displayed. Integrating displays which will simplify the execution of curved or segmented paths made possible by MLS will be in the new cockpit. We will see a major trend to a master navigation computer with multiple sensor inputs from such systems as INS, OMEGA, VOR/DME, DME/DME, MLS and perhaps GPS. It is possible that full-time displays of nearby traffic of concern will find their way into airline and perhaps other cockpits, based on either air-derived or ground-derived information. They may be used for a variety of purposes ranging from simple monitoring of the performance of an automated ground system, to confidence-building for aircraft on nearby parallel approach paths, to potentially more sophisticated capabilities such as sequencing, station-keeping, etc.

The 1990 era transport cockpit will likely be an all digital, all electronic flight deck with high resolution multifunction displays that will monitor flight operations.

Automated systems and electronic displays will simplify aircraft procedures in the NAS. To streamline the control process, both ground-based and air-derived surveillance information, meteorological data, knowledge of intent, and other information will be exchanges between automatic airborne and ground systems. Both pilot and controller will monitor and assess system performance, becoming managers of highly automated systems. The NAS will offer high reliability and yet retain reversion capability to a backup manual control system. New systems in the future cockpits will include aircraft automatic control, fly-by-wire, flight management systems, advanced navigation systems, integrated multifunction displays, and digital avionics controls.

The high end of the General Aviation spectrum of aircraft will no doubt see the incorporation of multifunction CRT displays, flight management systems, advanced navigation and communications systems including data link capability. Cockpits will also include coupled autopilots, auto throttles, and an integrated EHSI to include a navigation display option overlaid with color radar weather data. Features such as on screen touch control, head up display, interactive voice response system and direct access communications between airborne and ground computers are currently in various stages of development.

On the low end of the spectrum, the degree of sophistication will be highly dependent on the owner's economic capability and operational requirements, not on the technology required to perform the functions. An experimental system now in the early stages of development within the FAA will provide the small aircraft with a real-time ground weather radar picture in the cockpit. This uplink capability can also provide NOTAM information as well as SIGMETS, AIRMETS, PIREPS, area weather forecasts, hourly sequency reports and ATC information. As with all other systems, the display capability will be based solely on the economic ability of the owner. It is estimated that a low cost, perhaps as little as $500, CRT display could be developed.

BACKGROUND

In 1975, a special DOT task force study in the FAA safety mission recommended that "FAA undertake a major safety research program to assure that future systems are designed around reasonable criteria for human error." Consequently, the FAA Office of Systems Engineering Management undertook a study to identify human factors problems associated with both air carrier and general aviation accidents and incidents. This FAA study entitled, "Program for Optimizing Crew Performance and Minimizing Human Error in Aircraft Cockpits (Response to DOT Safety Recommendation 10)" used as inputs safety statistics from a variety of sources and solicited the views of the aviation community for its perception of human factors problems and potential solutions. After a great deal of internal and external discussion, several major problem areas were identified as primary candidates for expanded effort and formed the basis for establishing our human factors program.

While research and development in human factors had been carried on for many years in association with specific projects, FAA determined in 1977 that a common thread existed between the programs and problems and that central management was needed to assure a fully cohesive program which responded to identified problems. The programs were grouped into two broad areas related to pilot and controller problems. Because even the term "human factors" is frequently misunderstood, we chose to talk about our programs in terms of the intended result; namely, Aircrew Performance Enhancement and Error Reduction (APEER) and Controller Performance Enhancement and Error Reduction (CPEER).

Today, I would like to give a brief overview of some of the efforts we have underway on the aircrew side. Many of these programs represent joint efforts between the FAA, NASA, and the Department of Defense, which were undertaken to assure that the Nation's best resources are applied efficiently to this problem.

AIRCRAFT COCKPIT AND AIRCREW HUMAN FACTORS ACTIVITIES

Our program in the aircraft cockpit and aircrew area consists of several types of activity including Human Factors Safety Analysis, Aircraft Certification, Airman Certification, and Aircrew/ATC Interaction.

HUMAN FACTORS SAFETY. In the area of safety analysis, we have established activities designed to quantify the problems and identify needed engineering and development activities.

o Pilot Error Analysis. Historically, pilot error is cited as a factor in approximately 60 percent of air carrier and 88 percent of general aviation fatal accidents. Pilot error is also cited as a significant factor in aviation incidents. A continuing study is being made of the types and causes of human error to establish a basis for improvement of current systems and design of new systems.

o We have underway a general aviation accident problem analysis. A detailed categorization of accidents attributed to pilot error and the identification of underlying human factors problems was needed. This was accomplished through a review of general aviation accident and incident data bases to determine and prioritize human factors problem areas. A significant part of this work included an examination of the relationships between weather-related accidents and current methods of instrument flight training. The program will also define and examine the effectiveness of alternative programs.

- <u>Commuter/Air Taxi Human Factors Problems Analysis</u>. Commuters and air taxis have the highest accident rate in commercial aviation, and our objective in this problems analysis is to provide the FAA operating services, the industry, and the engineering and development community with a better appreciation of the causes and priorities of human error problems peculiar to commuter operations.

AIRCREW/ATC INTERACTION. The addition of new systems, mentioned above, into the cockpit and ATC system has a complex effect on the controller/aircrew interface. We have developed several programs specifically intended to determine the liabilities and benefits of various methods of implementing advanced systems.

- Cockpit Data Information Requirements and Analysis. The introduction of advanced cockpit design concepts and advanced

ATC system improvements will present new requirements for cockpit information processing and display. It is essential that human and aircraft system capabilities work in harmony with the evolving ATC system. We plan to develop a series of recommendations for efficient means of displaying and using information in the cockpit, for consolidation of information on electronic displays and for functional integration of aircraft systems. Proper integration of such new capabilities as collision avoidance advisories, wind shear information, Microwave Landing System flexible approach paths, Cockpit Displays of Traffic Information, flight management computers, and others, is essential. A similar review of information requirements is planned for the helicopter area.

o Separation/Navigation Standards for En Route Operations. Since flight technical error of various kinds has a major impact on acceptable separation standards, this effort is intended to develop improved methods of analysis to guide determination of future separation standards with particular emphasis on the reduction of flight technical navigation errors of various causes. Our present program is examining the relationship between separation standards and navigation system performance for en route operations. Human error and blunders in navigation are signficant contributors to the failure of aircraft to navigate within designated routes. The program addresses the human factors problems related to the use of current VOR and area navigation systems which may contribute to the error and blunder problem. Important objectives are to examine advanced navigation system concepts and to establish the data base needed to define guidelines and criteria that will recognize the special needs of single pilot IFR operations.

- Cockpit Display of Traffic Information (CDTI). While the technology to provide traffic information in the cockpit exists, the pilots ability to use this information and the impact this will have on the ATC system is not fully known. Our objective is to evaluate the use of Cockpit Display of Traffic Information for both passive monitoring and active spacing tasks so that the advantages and disadvantages of such use can be measured in terms of system safety, capacity, and efficiency in operationally realistic environments. We want to evaluate the impact of CDTI on the pilot and on the controller, as well as the impact of CDTI on traffic flow stability, dynamic merging and spacing, display content and format, and pilot/controller workload changes. Closely related to this work are efforts to develop and evaluate optimal displays for the Traffic Alert and Collision Avoidance System (TCAS) and the Automatic Traffic Advisory Service (ATAS).

AIRMAN CERTIFICATION. These projects were developed in direct response to specific Aviation Standards requests for the particular research to be accomplished. They include investigations of training programs for private, instrument and multiengine ratings and proficiency training programs with emphasis on low cost mechanisms and the value of improving decisionmaking training.

- Private Pilot Certification. Since the cost and complexity of operating aircraft is increasing dramatically, this effort is to determine optimum training and proficiency check techniques for private pilot certification.
- Low-Cost Proficiency Training. In a related effort, since the cost of flying has risen dramatically, we are looking at low-cost methods to train and maintain pilot proficiency by increased use of simulation, although we are aware that there are real limitations in extensive use of simulators.
- Instrument Rating Study. Since most weather-related accidents occur with VFR pilots having less than 200 hours total time, this work is intended to shed light on the feasibility of granting an instrument rating certification prior to reaching 200 hours total time.

- We are conducting a study on the relationship of general aviation pilot judgment and training to aircraft accidents. Inappropriate judgment is suspected of being a prime cause of pilot error in general aviation accidents. Our objective is to develop a system of experiments to assess pilot judgment in selecting appropriate actions under varying cockpit, ATC and aircraft emergency-conditions. We plan to examine the feasibility of developing training for improved decisionmaking to determine if pilot judgment training can offer specific benefits. This program will result in development of a syllabus which will help teach proper decisionmaking, defined as the optimal application of learned facts, and validation of this concepts through field tests.

HUMAN FACTORS GUIDELINES FOR AIRCRAFT CERTIFICATION. The majority of our aircrew human factors efforts deal with the introduction of new systems and procedures into the cockpit and our support for the operating services in their certification processes. This includes the review of current regulations and procedures related to the human factors area with a view toward identifying potential changes related to desired system improvements.

- FAR Part 25 Cockpit Standardization. Transport aircraft have been examined with regard to cockpit standardization, to identify the potential problems that may relate to equipment non-standardization. We have conducted a survey of seven representative airlines to determine the present status of cockpit standardization between aircraft of the same type and between aircraft of different types as an aid to identifying any problems associated with non-standardization. The product of this work entitled, "Transport Aircraft Cockpit Standardization" reports information on the current status of standardization and the benefits of additional standardization and is currently being reviewed for comment by the SAE S7 Committee to assess the possible value of additional standards for commercial aircraft cockpits.

- <u>Altitude Callouts During ASR Approaches</u>. A program has recently been completed on the effect on pilot performance of controller altitude callouts during Airport Surveillance Radar (ASR) approaches. This program addressed the value of providing mandatory altitude callouts by controllers during ASR approaches in reducing landing accidents. The conclusion of this work indicated that the presence of altitude callouts did not significantly affect pilot performance in executing ASR approaches.

- <u>Aircraft Alerting and Warning Systems</u>. Current transport aircraft systems are being examined to determine those factors which could contribute to pilot judgment error and incorrect remedial actions. Further, current systems may not indicate the optimal order in which critical actions should be taken when multiple or catastrophic failures occur.

This program has been underway for several years, with participation from the three major U.S. civil transport aircraft manufacturers. Our objective is to develop guidance for the functional standardization of air transport cockpit alerting systems, particularly with regard to the use of automation and new displays of alerting and warning data. We have encouraged the airframe manufacturers to work together to coordinate the development of a standardized industry alerting system concept. A major study entitled, "Aircraft Alerting Systems Standardization Study" has been completed which lays out the dimensions of the problem, and recommends specific alerting system concepts which were tested in simulation. We are planning to go beyond this effort to concentrate on more advanced methods of warning which take into account the changing priorities for warnings with flight phase.

o Computer Aided Decision Making. We have been working on research into more intelligent warning systems which can provide not only prioritized alerts and warnings, but which may also be able to provide diagnostic capabilities that will offer the pilot the best alternative course of action instantly, based on computer-aided analysis of the aircraft state. Current air carrier aircraft have complex emergency/failure procedures and checklists, and in cases of multiple system failures, the likelihood of intermingling checklist procedures is high and the consequences potentially severe. In this program we have investigated the feasibility of applying computer-aided decisionmaking to analyze complex and interacting aircraft systems so that unusual failure situations can be detected and remedial actions recommended to the pilot. We believe this work may show that computers having a knowledge data base and programmed reasoning ability can assist the pilot in high workload situations.

- Development and evaluation of Head-Up Display presentations for civil aviation aircraft has been undertaken. The program seeks to define alternative display presentations and assess the potential benefits and also any liabilities of this type of information presentation, in contributing to safer operations in air carrier aircraft during approach and landing. FAA has established a joint program with NASA to examine the potential of Head-Up Displays to aid the flight crew in reducing pilot workload, increasing reliability, and providing redundancy of information for navigation, flight path control, and other flight management tasks. The performance of flight crews using the device will be assessed over a full range of operational and weather scenarios. Our purpose is to provide enough basic data to the industry and to FAA's Aviation Standards organization to establish the capabilities, limitations, and minimum requirements for such systems.

- Our work on the wind shear program, which is essentially complete, included a great deal of emphasis on the human factors aspects of the problem; namely, how best to determine and present the information to the pilot. The airbone wind shear program began with a series of manned flight simulation experiments to identify and then refine the most effective pilot aiding concepts. Most subject pilots favored a system that displayed an airspeed-ground speed comparison. Another system that rated well in the evaluation utilized a "quickened" flight director logic. These results were validated in a number of simulations with airline and FAA pilots, and the results have been made available to the industry. The current effort is to determine and validate a wind shear model for use as a standard in simulator training.

- Alternative Collision Avoidance System Displays. Since threat alert and collision avoidance systems are new, this effort is intended to determine the minimum display requirements and to determine the optimum display characteristics and information requirements for such systems.

- General Aviation Weather Products. This effort is intended to demonstrate the utility of transmitting and displaying the low-cost weather information to light general aviation aircraft in order to help reduce weather-related accidents and incidents to non-radar equipped aircraft.

In addition to the programs mentioned above which have applications to helicopters, we also have underway or are planning a number of programs that relate specifically to the human factors problems associated with helicopters. One of these programs is designed to define the minimum acceptable handling qualities for IFR flight in helicopters. Other efforts include analysis of accident data and a survey of helicopter operators to identify potential helicopter problems and characteristics which may contribute to helicopter accidents.

o <u>Pilot Workload Measures</u>. Our effort here is to develop and validate workload measures to assist in certification of aircraft and new equipment. Although a great deal of work has been done on the subject of defining pilot workload measures, additional efforts are needed to develop fully acceptable, scientifically validated and widely accepted methods for measuring pilot workload. Some of the current efforts underway to deal with this problem include:

1. Completion of a report entitled, "Flight Crewmember Workload Evaluation" covering workload measurement techniques that have contributed to successful air transport certification programs.

2. A joint activity with United States Air Force to survey and categorize all existing or planned workload assessment and measurement techniques.

3. An effort to develop and validate a set of subjective pilot workload measures that can be used to assess reliably the workload associated with advanced cockpits of aircraft operating in current and future ATC systems. The intended end product will be a set of pilot rating scales for total workload measuement which is widely accepted and which can be used by government and industry researchers as a common measurement standard. As an initial activity in this program, the subjective workload rating scale developed by MIT, which is based on an earlier method developed by Cooper and Harper of NASA, is currently being examined and validated at the FAA Technical Center as part of the "Technical Assessment of Pilot Effectiveness (TAPE)" program. John Fabry from the Technical Center will present his recent research results on this, at this workshop.

4. Another approach being followed recognizes the importance of full mission system simulation in characterizing workload scientifically. FAA and NASA are working together on the development of such simulations to be used as an aid in learning more about establishment of objective pilot workload measures to augment the large body of empirical and subjective information which now exists. Full mission system simulation techniques will also permit improved studies of the interface between the pilot and the ATC system where many human errors originate.

SUMMARY

During the next decade FAA plans to conduct, with substantial support from NASA, DOD, industry and the universities, a comprehensive human factors program which will address the causes and impact of human error on aircraft and ATC system operation, and establish criteria and guidelines for future automated system designs which will minimize the occurrence and impact of human error.

The program will reexamine the allocation of separation assurance and spacing functions between the cockpit crew and the controller, and between the automated systems (both in the aircraft and in the ground ATC system) and the human operators (controllers and pilots.) Since the human operator problems and human involvement in system design and operation are the major thrust of this activity, we have established the program to assure that all elements of the aviation community have a voice in determining our program. Perhaps most important to the activities of this group is the examination of the programs underway to assess the impact of workload on the total system operation and effectiveness. I invite you to listen carefully to John Fabry the day after tomorrow. For, I believe, his Techniques for Assessment of Pilot Effectiveness will have a major impact on our collective ability to assess the impact of new systems introduction.

Historical Foundations of the AFAMRL WORKLOAD PROGRAM

Robert D. O'Donnell, Colonel, USAF
Wright-Patterson Air Force Base

Historically, the concept of workload has been a morass in which millions of dollars and many professional reputations have been lost. This has occurred in part because workload is a deceptively simple concept which people easily intuit, but cannot easily define. Emanuel Donchin has described workload as being very like pornography. Everyone knows what it is when they see it, but it is extremely difficult to get a definition which is generally accepted. In fact, one can find at least 20 distinguishable definitions in the literature, and probably 100 techniques which have been proposed at various times to measure workload.

In spite of these problems, the term workload will simply not go away. Its persistence argues persuasively that it must have at least some potential value in understanding human behavior. The reasons for this persistence relate to the massive technological changes introduced in the last several years. The technoloy being used in today's aircraft systems has introduced such complexity into the aerospace environment that the human is becoming the limiting factor in systems design. Whereas, in the past, physiological limits imposed by G forces, pressure, and other external stresses limited the aerosoace vehicle, it is becoming increasingly frequent to find that the human's cognitive limits now restrict potential applications. It is therefore essential that ways be developed to evaluate the load that existing and planned systems impose on the operator. Only then will designers be in a position to avoid dangerous overloads, design for optimal loading, and develop new procedures for expanding and superseding the human limits.

Recognizing this need, the Workload and Ergonomics Branch of the Air Force Aerospace Medical Research Laboratory, Wright-Patterson AFB, was formed in 1979. Its mandate was to establish allowable standards of workload in all Air Force systems, as well as to develop metrics to show compliance with such standards. As part of this effort, this Branch has surveyed the history of workload from the viewpoint of definitions and techniques and a general conceptual framework has been evolved for the area which hopefully can provide an initial foundation and structure for the definition of standards and the development of metrics.

"Definition" of Workload

As an initial, preliminary, and certainly incomplete "definition" of workload, the following statement encompases most of the concepts held by leading theorists in this field. <u>Workload is an hypothetical construct which conveniently describes the interactions between multiple factors affecting the operator's response in an operational system</u>.

It is important to recognize that workload is an hypothetical construct which, in effect, has no concrete existence. It has value only insofar as it <u>conveniently</u> summarizes many diverse things under one concept. These "diverse things" are those multiple factors which affect the operator's response. They are not, however, synonymous with the response. Rather, <u>interactions</u> between factors produce an overall effect on the operator and on system performance.

Obviously, such a broad and incomplete definition has value only if the factors underlying them can be identified, and if metrics to assess these factors can be specified. It is convenient to identify three broad categories of factors which contribute to workload, and then to break these broad categories into subcategories. Hopefully, measurement techniques can then be identified which will map onto each of these subcategories in such a way as to provide an overall assessment of the construct. In the categories to follow, the major classifications proposed by Johansen are utilized. These, in themselves, were evolved from several other theorists (Morey, 1979). They consist of: (1) task load, (2) operator variables, and (3) the response. In the sections to follow, each of these will be considered separately.

Task Load

Task load is defined simply as all of the behaviors required to achieve the task goal. It is extremely important to begin with this aspect of the workload construct, since task goal and the behaviors required to achieve it provide the reason for existence of workload in the first place. It can be seen therefore that workload analysis must start with the operational goal or mission. This demands that there must be a precise definition of what is required to achieve that mission. Without such definition, it is virtually impossible to approach questions of workload in any but the most abstract theoretical sense.

Specification of the task load involves at least three major subcategories. Perhaps the most fundamental of these, the <u>sensory environment</u>, is the one which is most frequently overlooked. The temperature, humidity, vibration, g-forces, etc., in which a behavior occurs significantly affect the difficulty of the task. Workload measurement procedures have seldom taken such factors into account, and no established methodology for including this type of influence in the overall workload equation has been developed.

A second subfactor of the mission which must be considered is the cognitive demand which can be anticipated in the task. Cognitive demand is defined as those internal operations necessary to achieve the task demands, and include things such as memory, decisions, and other mental operations which are obviously required of the operator. As aerospace systems evolve, there is a clear trend toward an increase in the cognitive demands of tasks at the expense of manual control tasks. Yet, few mission description techniques are capable of handling such cognitive demands, or even clearly specifying which cognitive operations are required to achieve a given goal.

The final and most obvious task load subfactor involves motor behaviors. These consist of those things which must be done to the outside environment in order to achieve the task goals (such as control behaviors, switch activations, etc.) and system dynamics which impose requirements on the operator to carryout specific error-nulling responses. Perhaps the greatest amount of work with respect to aircraft systems and workload has been done in one aspect of motor behavior, namely, manual control. Elegant models can permit rather precise definition of the control factors critical to a closed-loop, continous system. Less well specified, but still extensively studied, have been the discrete performance tasks required from the operator. Data bases consisting of task completion times and, in some cases, variability estimates are available. These can be incorporated into overall mission task analyses.

In many cases, task loading, especially the motor subfactor, can be assessed by task analyses, time lines, and by various systems models which have been used over the years. These analyses and models essentially ask whether the job can be done. More simply, they ask if there is enough time to do all that has to be done. As such, task analyses are critical to any workload analysis and constitute the starting point in deciding whether the system is even feasible.

However, in many cases, task analyses and systems models have been viewed as the final workload measure, and have been presented as the definitive evidence that a system does or does not meet workload standards. Aircraft manufacturers have relied heavily on these procedures to demonstrate to the FAA that a proposed system is certifiable with respect to workload. Even ignoring the fact that such models have obvious weakness in the area of sensory and cognitive factors. Task analyses almost always fail to consider the operator as a major factor in system operation. Such failure can result in a system which, while flyable under optimal human conditions, is not designed for the real operational enviornment. In that environment, the operator is a major determinant of the task demands as well as the major mechanism on which those demands act.

To present this more forcefully, Figure 1, shows that there is a two-way interaction between the task demands and the operator. In effect, the task demands and the operator together form the person-machine system. Any workload metrics must consider this system as a whole. No existing model adequately achieves this goal and, at best, such models or task analyses statistically approximate the mutual interaction between task demands and operator. Since little is known about whether such interactions are additive, multiplicative, or much more complex, such approximations are always crude.

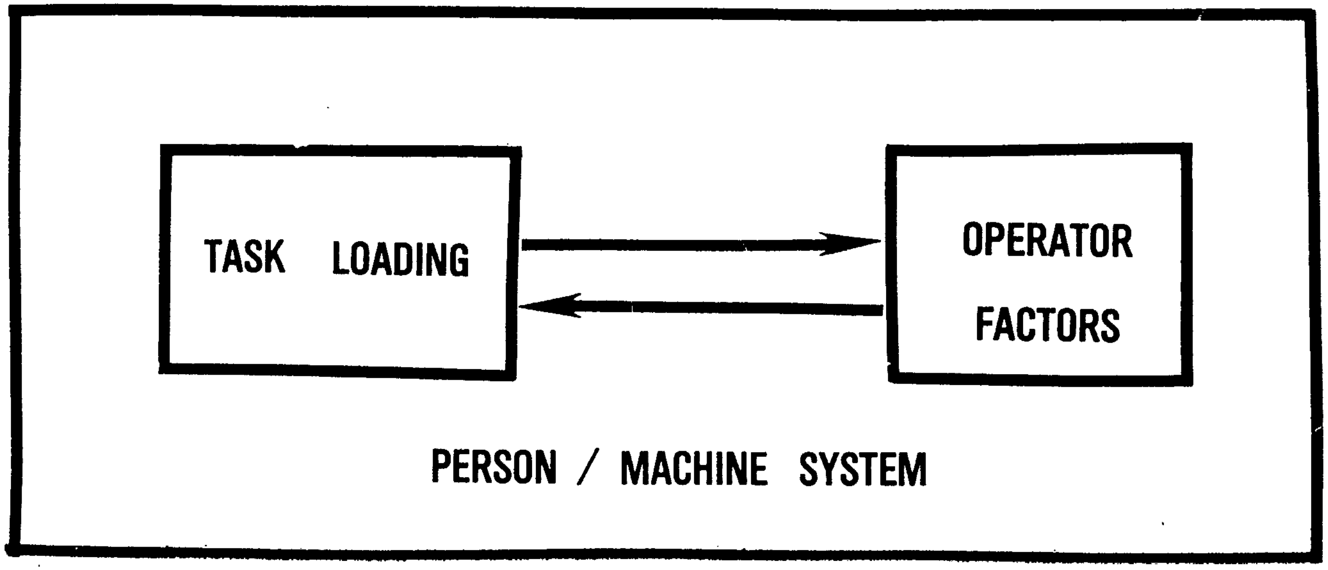


FIGURE 1

Operator Factors

If the operator must be considered as part of the overall system, those subfactors constituting this major influence must be identified. Clearly, the operator's physical capacity is a basic subfactor in this category. Although in most aircraft environments adequate selection makes this a less crucial variable, it assumes considerably greater importance if the operator may be decremented due to injury or toxic exposure. In such cases, workload analysis must consider the interactions generated by a less than optimal physical condition. Analogously, mental capacity must be evaluated. Again, while selection in most aircraft enviornments assures adequate initial mental capacity, long-term changes due to stress, fatigue or other decrementing conditions must be anticipated. Similarly, the subfactor of emotional condition must be specified with respect to the workload environment. These three categories of physical, mental, and emotional state have not usually been considered necessary in workload assessment. Yet, in many cases they constitute the questions of greatest interest to the operators themselves, who are aware of transient changes in these factors and wish to be reassured that the workload can be handled under decremented conditions.

The most obvious operator subfactor to be evaluated is the effect of task demands on the operator's mental resources. Given adequate physical, mental, and emotional capacity, there is still the question of the short and long term effects of performance.in the enviornment on the operator's reserve mental capacity, on the strategies employed in meeting the task demands, and on questions of attention and resource allocation over time. Simply stated, an aircraft's tasks may be do-able by the rested, trained, motivated pilot, but they may not leave any margin for safety, and sustained performance over a period of time may significantly affect what margin existed originally. For such questions of mental resources, whole series of procedures have been developed. Briefly, these procedures can be viewed as falling into three major classes.

Subjective measures are those in which the operator is simply asked to evaluate either his or her own workload or that of another operator. These measures are based on the assumption that the human being is the best observer of all the interactions between factors entering into the workload construct. Many types of subjective measures have been used, including modifications of the familiar Cooper-Harper scale, as well as newer, advanced techniques such as conjoint analysis.

Many believe that subjective assessment techniques are the best single measure of workload. Certainly they have good face validity and have produced the best results in the past. One must be aware, however, of their limits. Subjective techniques may not be very helpful in some situations, particularly where the purpose is to diagnose the cause of the workload problem. Although pilots may frequently be able to elaborate on a subjective rating and to provide useful insights into the source of the workload problem, it is not always desirable to place heavy reliance on such analyses. Individual motivations, capacities, or preference may bias results and can lead to false conclusions (for instance, the "old boy syndrome" may perpetuate poor but familiar procedures). In spite of this limitation, however, as Sheridan and Jex have maintained, subjective measures will be extremely important for a long time to come, and considerable efforts should be expended in improving these techniques.

Behavioral measures consist of two major types. Primary task measures assess the person's performance within the system. More will be said about this type of measurement later. Secondary task measures involve the imposition of an additional task on the operator in order to "load" the person to maximum capacity. The break point in primary task performance can be used as a measure of workload, or the secondary task can be used to provide an independent assessment in its own right. Many types of secondary tasks are available. The workload assessment device (WAD) currently used at Patuxent River, Maryland by the Navy was initially developed at AFAMRL, and represents a good example of this type of technique currently being used in operational environments with great success.

At AFAMRL, primary emphasis has been on the development of non-intrusive secondary tasks. In particular, one type of task has generated considerable interest. This has been called the embedded secondary task. In this procedure, a part of the primary task (such as the communications subtask in a flying task) is taken into the laboratory and quantified with respect to workload. Once the workload involved in this subtask is well defined, it can then be reintroduced into the primary task. As far as the operator is concerned, no artificial task has been introduced, and yet a quantified secondary task is being presented, thus yielding the advantages of the secondary task methodology. This procedure has been developed for the A-10 aircraft communications task, and is currently being tested in a simulator environment. Current plans call for it to be available in the near future for field use.

The final technique for assessing operator variables consist of neuropsychological procedures such as the cortical evoked response, heart rate measures, muscle measurements, blink rate and eye movements, etc. which can measure the effect of task load directly on the person. It is important to recognize that these procedures are not measuring physiology in any real sense, nor are they looking at the medical condition of the pilot. Rather, they look at the brain, muscle, and heart responses which are given by the operator. In this sense, they are no different than a verbal report or a motor response. However, they are obtained more quickly and, hopefully, more reliably. We are currently testing a neuropsychological test battery which contains eleven such standardized procedures. These are being validated in Air Force environments, and we hope to have at least some of them available for flight test in the near future.

With these subjective, behavioral, and neuropsychological procedures, we hope to assess the operator's interaction with the task load more precisely, and to provide the data base necessary to permit adequate modeling of the person-machine interaction.

Response Factors

Even when the above goal has been achieved, there is still another major factor influencing the overall workload equation which must be considered The response is really the major concern in Air Force systems. Certainly, one would expect that the response will be correlated with the workload imposed on a person-machine system. In fact, many would propose that system performance (primary task performance) be made the ultimate workload measure. This argument says simply that if there is no decrement in system performance, workload is acceptable. The problem with this logic is that the response is not always manifested in short-term system performance. In many situations, the person performs perfectly until catastrophic failure. For this reason, it is more precise to subdivide the response into two elements. System performance, as defined above, is one of those elements. The other is the "cost" to the operator. This is defined as the decrement to an operator's capacity as a function of maintaining performance. While the workload of the person-machine system determines the response, this response can be manifested either in a performance effect or in an effect on an operator's capability. This, in turn feeds back into the system, as shown in Figure 2. If performance is affected, it affects the tasks to be done in order to achieve the mission goal. On the other hand, if the person's capacity changes it affects the operator's factors. In either case one can conceptualize essentially all of the various interactions constituting the workload construct within this dynamic framework.

Implications for Workload Assessment

Using the above framework, several principles of workload assessment appear evident. First of all, it should be clear that a measure of any one of the factors shown in Figure 2 is not in itself a workload measure, unless all other factors and interactions are specified or controlled. In other words, no single factor is the workload. The confusing and frequently contradictory results found in the workload literature have

WORKLOAD FRAMEWORK

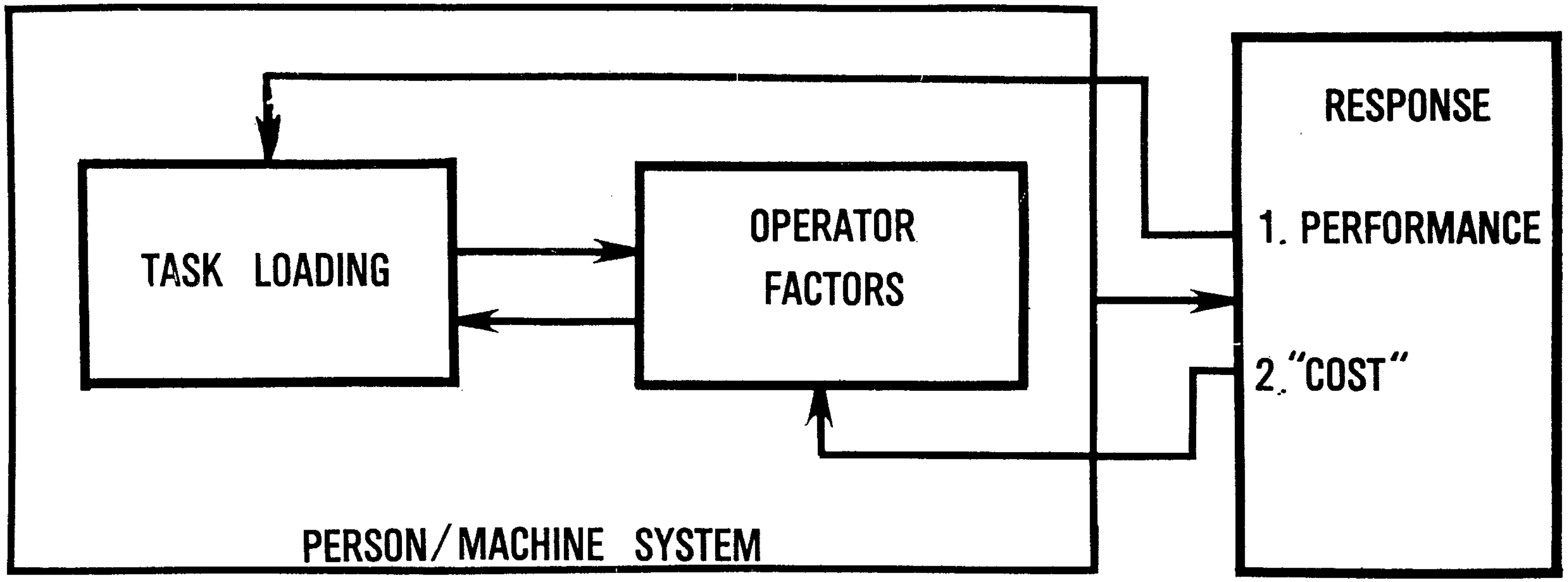


FIGURE 2

most often ben due to the fact that individuals have measured one factor while failing to control or account for all the others. The practical impact of this principle is that a workload program must be multidimensional, and must include techniques for approaching all of the factors involved. In addition, any attempt to measure workload in a system must give considerable attention to discussing potential interactions between factors which are not measured in that particular study. It is possible, for example, to do a perfectly adequate workload assessment without considering decremented environmental conditions. However, in such a case, it is necessary to clearly define the limits of environmental conditions to which any results would be applicable. Awareness of such factors would preclude a great deal of overinterpretation of workload studies, while still permitting valuable data to be collected upon which decisions can be made.

A second implication of the above framework is that the search for a "holy grail" of workload assessment should probably be abandoned. No single measure, including subjective assessment, will ever be able to adequately assess all of the multiple factors and interactions. To this end, the AFAMRL Workload Program has dedicated itself to development of a broad range of metrics which can be included in test batteries. In this way, the historical lessons learned from past attempts to assess workload can be applied on a case-by-case bases to specific workload questions. Tests can be chosen from the batteries which answer the question being asked, rather than attempting to answer amorphous general workload questions, or answering a question which was not asked in the first place, simply because a measure is available.

In conclusion, nothing that has been discovered about workload in the last ten years has changed its basically complex nature, nor has it revealed a simplistic way of measuring it. However, significant progress has been made in defining what the various critical elments of workload are, and in pointing out those elements which can be measured at the present time, as well as those which will require new metrics for their measurement. At AFAMRL, such metrics are being developed, and it is hoped that they will be applicable to all environments. There is considerable optimism that such a goal is achievable, and we intend to achieve it.

Pilot Models for Flight Control Analysis

R. O. Anderson

Flight Dynamics Laboratory

Summary: After a brief definition of "pilot models", the two most common models ("classical" and "optimal") are briefly described. The basic limitations and advantages of each are summarized. Next, some of the potential uses of these models are discussed along with a few non-aircraft oriented applications. The available data base to confirm models is reviewed and it is concluded that despite many very successful applications to date, overall understanding, use, and actual in-flight data confirmations are limited.

Pilot Models: As used here, a "pilot model" is some form of mathematical expression that attempts to portray how the human operator behaves in certain "control" situations. Perhaps the first crude model was in N. Minorsky's mind when he wrote, in 1922 (Ref. 1)

> "It has often been stated that the human intuition of the helmsman cannot be replaced by any mechanical contrivance whatever its mature may be. Such a standpoint seems to be erroneous."

He then proceeds to describe the human operator (helmsman) action as responding to ship yaw plus yaw rate. In short, a verbal description of the helmsman as a linear control system that we could represent with a differential equation, or "pilot model".

No attempt will be made here to trace the detailed history of mathematical models of human behavior; instead, the reader should consult Refs. 2 and 19 for such a discussion. The purpose of this brief review is to emphasize a few milestones in the story and, if possible, bring the current status into perspective, at least from the author's viewpoint.

Classical Servo Model: Dubbed "classical" this model was the earliest to gain widespread credibility. It simply assumes the pilot's input-output response for certain tasks is similar to a "good servo" system. This reversal of purpose from Minorsky (who wanted to build a "servo" that operated like his "model" of the human) was made possible by some 35 years of linear control theory development. Now, the classical model is simply a fundamental application of undergraduate linear control theory.

The tasks mentioned above must, however, reduce the human operator to somewhat of a "monkey" status. That is, the parameters to be controlled must be well known (e.g., hold aircraft pitch attitude), the inputs or disturbances must be "random appearing", the operator must devote full attention to the task, and the pilot/aircraft system must represent an equivalent single-input and single-output linear compensatory system. Work has been done to relax every one of these restrictions with some varying degrees of success, but for the most part, they must all hold for a "confident" application.

The good news for modelers is the above "monkey" restriction doesn't seem to preclude application to the task of flying a real aircraft. This is also the bad news for pilots.

Figure 1 shows the heart of the classical model. With the pilot quasi-linear describing function given as Y_p, and the aircraft output (m) to input (c) transfer function fiven as Y_c, the name of the game is for the pilot to vary his response to the error signal (e) such that the product Y_pY_c becomes that of a pure integrator (k/s) in the region of crossover. The latter is the frequency range where the Y_pY_c amplitude ratio becomes unity on a Bode plot. This "goal of adaption" is one version of what it takes to make a "good servo" (Ref. 2). Does it work? Yes indeed. In fact, the author has seen measurements that indicate this not only in the case for good old U. S. of A. pilots, but "trackers" from several other countries. If you still don't believe, check Ref. 3 which probably still represents the most carefully conducted and complete experimental work in the pilot model parameter identification area.

As an interesting aside, the experiments in Ref. 3 utilized a data reduction scheme using common utility company watt-hour meters because they were the most cost-accuracy effective analog computers at that time.

Despite numerous applications that many consider an unqualified success, the classical model still has its limitations. In short, only single-input, single-output tasks are fully understood and validated. The task must be well known and clearly defined, and must satisfy the "monkey" conditions. Nonetheless, the classical models have been applied to some extremely complex tasks.

On the other hand, the biggest advantage of the model are its simplicity and the fact that it works!

Optimal Pilot Model: Before proceeding with this discussion a word of caution is in order. While assisting another engineer at Wright-Patterson Air Force Base with a problem of evaluating turbulence effects on a new proposed aircraft payload, it was suggested that a "pilot model" be used to analytically control the unstable spiral mode of the aircraft. The engineer's reaction was that this could not be done without "picking one company's or the other's model", and this was a no-no. Unfortunatley, pilot models may have become associated with one person or company. Therefore, keep in mind the type we are discussing all assume the human operator acts like a "good servo". How you design the good servo is the key "individual" distinction between the model (see Ref. 19 for a comparison of the classical and optimal models).

Now, on to the optimal model. Few know that this now famous creature was born in the lounge of the old Boston Hilton. Seems BBN

(Ref. 4) was having trouble using the "classical" model under an Air Force contract to apply pilot models to aircraft display requirement determination. Viewing the pilot as an optimal controller had been suggested by the author (and other) in an IEEE Journal Note. Jerry Elkind, then of BBN, suggested this be attempted under the contract one night at an informal "conference". The answer was yes, but good luck selecting the performance index. That started it all.

The optimal control model, shown in Fig. 2 simply assumes the pilot has some mental "performance index", and being the "good servo" that he is, he minimizes the index. This index is a quadratic function of aircraft state (x) and control input ($\dot{u}$). Once the index is available, a common control theory "meat grinder" will produce the optimal controller which becomes the assumed pilot model. The net result is the same as the classical approach; namely, "prediction of closed-loop performance and pilot model parameters". Foes argue that no "real" pilot knows a "quadratic performance index" from a flame-out, that the results with a single performance index will match observations, etc. Friends say it's not a "real" model, but does predict performance and it handles multi-input, multi-output problems readily (the above meat-grinder is very powerful in this regard). Suffice it to say, it's another way to design a "servo" (pilot model), and it also works.

Other Models: Again, no attempt will be made to list the dozens of derivatives of the above two models that have been postulated over the years. Reference 2 contains references to many. Typical of these is the "dual-loop" model in Refs, 5 and 6, and the "step-tracking" model in Ref. 7. The interesting thing about both of these is that they are being used in an attempt to predict, or evaluate, flying qualities. This was, in fact, the noble goal of Charlie Westbrook in the AF Flight Control Laboratory back in 1954 when he sponsored much of the early work in pilot modeling that he hoped would see fulfillment in 5 or 10 years.

The details of model development, applications, and general engineering acceptance are too numerous to cover here. Reference 2 and the "Annual Manual" proceedings (e.g., Ref. 8) cover much of the story, both successes and setbacks (e.g., see Refs. 40, 44 - 54 in Ref. 2). The Annual Manual really started when the Air Force decided to have the few researchers in the area (sponsored by the AF and Navy) meet once a year to discuss progress. In fact, one meeting was adjourned to the author's house for beer, pizza, and further discussions. Now the annual meetings attract 75 to 100 people.

The cast in these developments is also too large for much individual acknowledgement. Therefore, at the risk of raising the wrath of those I leave out, the classical U. S. leaders that must be mentioned, and are still in the business, are: McRuer, Krendel, and Graham. Those moving on to other areas include: Tustin, Elkind, Russell, Sobczyk, Phillips, and Weiss.

The U. S. optimizers have held in a bit better with those still making a sound living on the "optimal servo": Baron, Levison, Kleinman, Hess, and most recently Dave Schmidt of Purdue. Fallen by the wayside are Elkind, Peter Falb (who extended modern control theory which at the time could not include the known pilot pure time delay), and Jim Dillow (Ref. 16) my partner in crime who made pilot modeling "fun".

Pilot Model Uses: It would be nice to say at this point that every high school graduate understands, and uses, pilot models. This just isn't so. While a large number of applicantions do exist (again see Ref. 2) they have been accomplished in the U. S. by only a few people. Fortunately in Europe (where a European Annual Manual was recently started) the situation seems to be different. In any event, a few applications to aircraft control are covered here.

Perhaps the most worked, and reworked, area of application is "human gun tracking". In fact, the early work of Tustin in 1947 (see Ref. 2) concerned this task. About the same time Phillips and Sobczyk also analyzed this task (Ref. 9) and even adjusted system parameters to minimize mean-squared-error; which Schmidt is doing today for the air-to-air gun tracking task (Ref. 10).

The next application area, prediction of flying qualities, motivated much of the early work in pilot modeling. However, the number of serious attempts to quantitatively predict numerical values of pilot acceptance (e.g., Cooper-Harper rating scale) are very few. In most cases, the basic concept is that pilot acceptance is degraded as closed-loop performance degrades and/or "workload" increases. The key issue is what constitutes a suitable measure of "performance" and "workload"?

Reference 11 represents my first contribution to the rating prediction confusion. Basically, the conjecture is that "performance" is some measure of closed-loop root-mean-square-error values, and "workload" is generated lead (value of lead time constant in the pilot model). The kicker is the pilot model parameters are adjusted, within measured limits, to minimize pilot rating (best overall acceptance).

The concept, called "Paper Pilot", is actually vwey simple, although the original digital computer program to implement this approach was fairly large and complex. This fact, along with an apparent virtue of the approach that backfired (i.e., completely automated flying qualities numerical prediction) may have hindered acceptance. The latter point may follow the story of the early non-acceptance of pre-mixed cake mixes, which prevailed until the concept was changed and the housewife was required to add an egg to the mix. With this "personal" touch, all went well.

Shortly after the Paper Pilot, Neal and Smith at Calspan (Ref. 12) proposed longitudinal "criteria" for flying qualities that also used a pilot model. Again, ratings were correlated with closed-loop performance (in this case it was damping ratio which is related to the root-mean-square values in the Paper Pilot) and pilot "workload" (i.e., pilot model lead, or lag, time constant with the former the same as Paper Pilot). The model parameters are "egged" into line by trial-and-error to achieve a closed-loop bandwidth with a limit on low frequency "droop" in the aircraft pitch transfer function Bode plot. The results are plotted against constant pilot rating contours to evaluate the configuration in question.

Although the Neal-Smith criterion appears to be receiving greater acceptance than the automated Paper Pilot (the former apparently has the "egg"), neither have made it to Spec-dom (Ref. 13), and use of pilot models for rating prediction remains an amature sport.

While the flying qualities community continues to rush into the twentieth century, the grand-daddy of uses, gun tracking performance improvement, may be coming into its own. Schmidt (Ref. 10) has applied the optimal model to the problem quite successfully, and has generated a great deal of renewed interest in this subject.

But, what about other applications. Well, aircraft deplay selection and location seems, or seemed, a natural and the optimal model came about from such an application (Ref. 4). Much more could be done in this area.

In a similar manner, controller design might be another application for extended models (Chapter IV of Ref. 2). At least one such case is known. A collegue in the Army called me about yet another gun tracking situation in a tank. The "pilot model" produced markedly better performance than real gunners. After checking their model and finding no obvious boo-boo's, I asked about the controller. It was a two-axis "coolie hat" on top of a vertical post operated by the gunner's thumb. Try it, it's awful! I told him to replace this with a good force stick and try again. He called a few weeks later and said now all was well. Wonder if they ever used the new controller in a real tank based on a phone call?

Another long overdue application is simulation validation. Closely related to the flying qualities use, if the pilot models can't fly it, the simulation is in trouble. I was asked to do this once in our large hybrid simulation facility. It was kind of weird using a half-dozen operational amplifiers to validate a several hundred op. amp/32K digital simulation!

A related incident dealt with in-flight simulation. Back in 1970 the Navy was very proud of an F-8D being operated as a variable-stability aircraft. I took the oldest set of flying qualities data for the longitudinal axis I could find that had pilot model "data" (1958), and created the Pitch Paper Pilot (Ref. 14). This model was then used to predict (validate) the latest F-8D results (Ref. 15). Some of the correlations were good, some bad. However, a pilot model reaction delay time of 0.42 seconds had to be used in the model to get even reasonable results, vs. a value of about 0.2 usually used when motion cues are present. About eight months later I learned the

Navy had "found" a pre-filter in the F-8D (I don't know who "lost" it). In turn, it had a 2.5 rad/sec break point which produced a phase lag that almost exactly accounted for the required 0.22 second increase in pilot model delay. Paper Pilot had "invalidated" an in-flight simulation with a model based on data from 12 years earlier!

Finally, we may someday reach Minorsky's goal and model the human on our way to forming a "true" adaptive flight control system. McRuer has often called the pilot the "vocal adaptive controller". This use is self-evident.

Many other non-aircraft applications are available, ranging from auto driver models to Minorsky's problem, the helmsman. The Netherlands has tried the latter (unsuccessfully) as well as bike rider (successfully) applications. Shopping cart driven models anyone?

Pilot Model Parameter Data: Sooner or later this long "short story" has to cover flight test if it is to be presented at a meeting at AFFTC. Now is the time.

Ground-based simulation data to support both the classical and optimal pilot models are quite readily available for single-loop tasks with simple controlled elements (e.g., Ref. 3). Many of the optimal cases include only closed-loop performance correlations, but a few convert the complex pilot model into Bode plot form for comparison with measured describing functions or other parameter data.

When it comes to "real" aircraft dynamics, the data bank dries up. However, we did conduct some pilot describing function experiments using the B-1 simulation at the Flight Dynamics Laboratory. The really unique aspect of the effort, however, was the fact we had help from a Navy Post Graduate School Professor and used an Army describing function analyzer!

As for pilot model parameters measured in-flight, they are few and far between. After some earlier initial attempts (e.g., Ref. 20) Newell (Ref. 17) obtained ground and in-flight simulation results for roll tracking. This feat was repeated in-flight somewhat later with the then new NASA Jet-Star. Shortly afterward, Wingrove (Ref. 18) reduced some actual data from a Gemini retro-rocket firing under manual attitude control. This one experiment almost put our space age data bank (1957-present) even with our aeronautics (1903-present) bank. Not very impressive, especially in light of modern instrumentation/data reduction improvements.

The bottom line of this section is obvious. The workshop for which this was written (AIAA Workshop, "Flight Testing to Identify Pilot Workload and Pilot Dynamics") is clearly long over-due in respect to pilot dynamics in general, and pilot parameter identification in particular

Conclusions: Despite what many (including the author) feel have been unqualified successful devlopments and applications of pilot models, only limited applications and understanding are really evident in the U. S. This was driven home to me at a recent AGARD Flight Mechanics Panel Symposium where one attendee, at a break, suggested that very few in the audience understood the last paper which was based on pilot models. When I asked him what we could do about the situation, he had no concrete advise. Perhaps this workshop, and this paper, will kindle some new interest and new practitioners. If they do, the field is still wide open. We need multiloop and in-flight pilot model parameter data, extensions of the models to tracking plus discrete decision tasks, and more example applications with actual aircraft comparisons to increase general understanding and acceptance.

As for the future, well Minorsky and Tustin, we're still trying, and we may have some additional help pretty soon.

References

1. Minorsky, N., "Directional Stability of Automatically Steered Bodies", Journal of the American Society of Naval Engineers, Vol. 34, 1922.

2. McRuer, D. T. and E. S. Krendel, "Mathematical Models of Human Pilot Behavior", AGARD-ograph No. 188, Jan 1974.

3. McRuer, D. T. et al., "Human Pilot Dynamics in Compensatory Systems: Theory, Models and Experiments with Controlled Elements and Forcing Function Variations", AFFDL-TR-65-15, July 1965.

4. Elkind, J. I. et al., "An Optimal Control Method for Predicting Control Characteristics and Display Requirements of Manned-Vehicle Systems", AFFDL-TR-67-187, June 1968.

5. Smith, R. H., "A Theory for Handling Qualities with Applications to MIL-F-8785B", AFFDL-TR-75-119, 1976.

6. Hess, R. A., "Dual-Loop Model of the Human Controller", AIAA Journal of Guidance and Control, July-Aug 1978.

7. Onstott, E. D., et al., "Prediction and Evaluation of Flying Qualities in Turbulance", AFFDL-TR-71-162, Feb 1972.

8. Second Annual NASA-University Conference on Manual Control, Washington, D.C., NASA SP-128, 1966.

9. James, H. M., N. B. Nichols, and R. S. Phillips, "Theory of Servo-mechanisms", New York, McGraw-Hill, 1947.

10. Schmidt, D. K., "Optimal Flight Control Synthesis via Pilot Modeling", AIAA Journal of Guidance and Control", Jul-Aug 1979.

11. Anderson, R. O., "A New Approach to the Specification and Evaluation of Flying Qualities", AFFDL-TR-69-120, June 1970.

12. Neal, T. P. and R. E. Smith, "An In-Flight Investigation to Develop Control System Design Criteria for Fighter Airplanes", Vol I, AFFDL-TR-70-74, Dec 1970.

13. "Military Specification, Flying Qualities of Piloted Airplanes", MIL-F-8785C, 5 Nov 1980.

14. Anderson, R. O., A. J. Connors, and J. D. Dillow, "Paper Pilot Ponders Pitch", AFFDL-FGC-TM-70-1, Nov 1970.

15. Kaster, T. M., J. A. Eney, and J. J. McCue, "Flight Evaluation of Various Short Period Dynamics in the Variable Stability F-8D", NADC-NATC Report No. FT-13R-70, March 1970.

References Continued

16. Dillow, J. D. and D. G. Picha, "Application of the Optimal Pilot Model to the Analysis of Aircraft Handling Qualities", Aug 1975.

17. Newell, F. D., "Human Transfer Characteristics in Flight and Ground Simulation for the Roll Tracking Task", AFFDL-TR-67-30, April 1968.

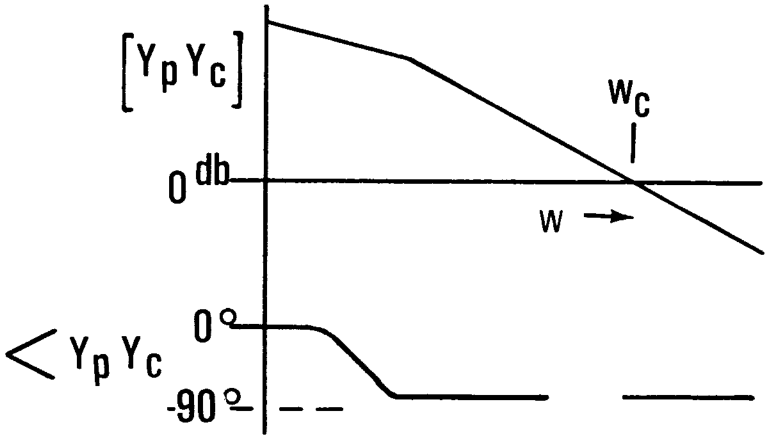
18. Wingrove, R. C. and F. C. Edwards, "Measurement of Pilot Describing Functions from Flight Test Data with Examples from Gemini X", NASA SP-192, 1968.

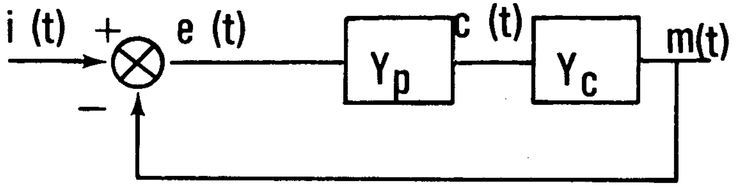
19. McRuer, D. T., "Human Dynamics in Man-Made Systems", Automatica, Vol. 16, pp 237-253.

20. Seckel, E, I. A. M. Hall, and D. T. McRuer, "Human Pilot Dynamic Response in Flight and Simulator", WADC TR-57-520, Aug 1958.

CLASSICAL MODEL

FDL FIG 1-2-5--4





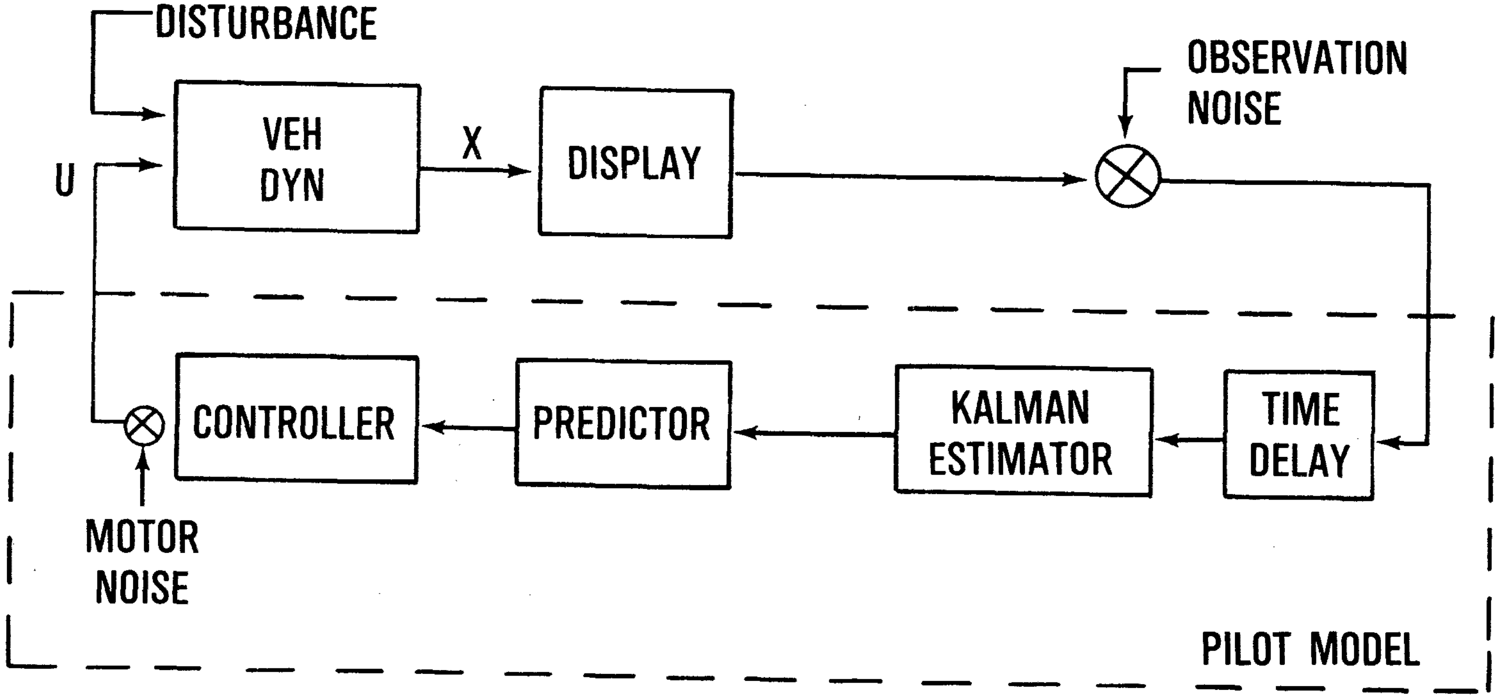
$$\frac{m\,(s)}{i\,(s)} = \frac{Y_p Y_c}{1 + Y_p Y_c}$$

Y_c	$Y_p Y_c$	τ_e
K_c	↑	—
K_c/s		0.14
$\frac{K_c}{S\,(S+\alpha)}$	$\frac{W_c e^{-\tau_e s}}{s}$	0.16
$\frac{K_c}{S^2}$	↓	0.43

G→

FIGURE 1

"OPTIMAL" PILOT MODEL



$$I = E \left\{ \frac{1}{T} \int_0^T (X' Q x + u' Ru) \, dt \right\}$$

KEY PLAYERS

J. ELKIND
P. FALB
S. BARON

D. KLEINMAN
W. LEVISON
R. HESS

D. SCHMIDT

FIGURE 2

NLR MP 82002 U

NLR RESEARCH ON PILOT DYNAMCIS AND WORKLOAD

by

R. C. van de Graaff

National Aerospace Laboratory NLR
Amsterdam, The Netherlands

INTRODUCTION

This paper reviews some results of NLR research on pilot performance and workload both in manual control and monitoring tasks. A substantial part of the research has been devoted to the description of human control behavior and task conditions in terms of modern control theory. The principal results of this modelling approach are considered in the first part of this paper. It is shown that the modelling approach can provide an adequate framework for analysis of pilot dynamics and workload.
The second approach, discussed in the subsequent part of this paper, deals with the assessment of pilot workload by means of physiological parameters. The results of an in-flight experiment reflect a good correlation between measures of heart rate and respiration frequency, model results, subjective ratings, overall performance and control activities.

MODELLING APPROACH

General

During the last decades considerable research effort has been devoted to the development of mathematical tools for the analysis of human control behavior. Among several different approaches towards human operator modelling the optimal control model has emerged as a most useful one for the study of complex manned aerospace systems. The model which is documented extensively in the literature (e.g. Refs. 1 and 2) is based on the assumption that the well-trained, well-motivated human operator behaves in a near optimal manner subject to the extent to which he understands the objectives of the task and his inherent limitations and constraints. The model inputs include task-related parameters (system dynamics, disturbance environment, available information) and human operator parameters (perceptual thresholds and time delays, human neuromotor noise, observation noise and the objectives of the task). The latter can be considered as the input variables of the model. The output parameters of the model include measures of system performance, human control activity and various attentional characteristics.

However, for a complete description of the human control behavior and its impact on overall system reliability, it is necessary to include the concept of workload within the optimal control model (OCM) framework ; i.e. to assess in terms of OCM parameters how hard the human controller has to work to achieve a given performance (criterion). In this context a control effort model has been developed at NLR (Ref. 3), which will be briefly reviewed in the following section.

Workload model

In the control effort model (Ref. 3) it is assumed that the human control response is partly determined by mechanisms that selectively tune the organism to the stimulus situation, by which is meant both selectively attending to some stimulus in preference to others and investing more or less attention per source of information. This can be identified with volontary attention (Ref. 4), reflecting that the subject attends to the stimulus because of its relevance for performing the task and not only because of its arousal function. Also involuntary attention is included in the control effort model. This can be related to the level of arousal and is largely dictated by the properties of the displayed information.

The aspect of voluntary attention is incorporated in the model in terms of the overall level of attention, P_o (following reference 6). The aspect of involuntary attention is included in the control effort model in terms of the sensitivity of task performance (cost functional, J) to the momentary attention paid by the subject. Now, the human operator's control effort (the workload index E) is defined as the ratio between the performance sensitivity S and the overall level of attention P_o.

In formula:

$$E = S/P_o \qquad (\mathrm{dB}) \qquad (1)$$

with

$$S = \frac{\partial J}{\partial P_o} \qquad (\mathrm{dB}) \qquad (2)$$

where the partial derivative indicates that the other model parameters are kept constant. (The cost functional, J, is the weighted sum of the pertinent system variables).

In this way the optimal control model has been extended by a task effort model so that the workload aspects of the task can also be assessed within the optimal control model framework.

Predictive capability

The usefulness of the model to describe human control dynamics and workload has been supported in various control tasks under laboratory conditions, as well as in realistic in-flight helicopter control tasks. This is summarized in the following ; for an extensive presenatation the reader is referred to references3 and 5.

The experimental variables in the laboratory experiment were chosen so as to include all possibly important characteristics of pilot behavior in control situations. Pilot equalization was experimentally varied by incorporating controlled element dynamics of the form k, k/s and k/s^2, requiring lag, "pure gain", and lead equalization, respectively. Furthermore, sufficient variation in task difficulty was realized by varying the instability level of the controlled elements. For the eight single-axis control tasks of the experimental program the computed control effort results were compared with subjective ratings (McDonnell rating ("demand") scale,see reference 7). The results, which are shown in figure 1, indicated an excellent correlation between both.

For the in-flight validation experiment (Ref. 5) an instrument hover task and two navigation tasks on an Alouette III helicopter were chosen. The hover task consisted of stabilizing the helicopter at 600 ft altitude with minimal horizontal (ground) speed. The two navigation tasks consisted of flying along a desired track with an indicated air speed of 60 kts at a prescribed height : 600 and 150 ft, respectively. The instrument information consisted of the three attitude angles (provided by a three-axis ADI), height deviation, horizontal velocity components (in case of the hover task and indicated by the flight director bars), and cross track deviation (in case of the navigation tasks and indicated by the vertical flight director bar).

The performance index J was determined from the recorded display parameters:

$$J_{hover} = (\text{RMS } h/h_L)^2 + (\text{RMS } v_h/v_{h_L})^2$$

$$J_{nav} = (\text{RMS } h/h_L)^2 + (\text{RMS } y/y_L)^2$$

Both performance indices incorporate the scores which were instructed to be minimized : height error (h) and horizontal speed (v_h) for the hover task and lateral deviation (y) for the navigation tasks. These scores are weighted by the corresponding display limits (index L) so that J is the sum of the mean-squared fractions of the full deviations. This criterion is analogous to the cost functional of the control model, where it is assumed that this criterion is minimized by the human operator.

The comparison of the control effort model predictions (in dB) with the subjective effort ratings of the participating pilots are presented in table 1. As the table shows, the model predictions indicate the hover task as the most difficult flight task, whereas both navigation tasks are of the same difficulty. These predictions are thus in good agreement with the results from the subjective effort ratings.

Furthermore, the comparison of the model predictions of the pilot's control behavior and task performance with the experimental results (Ref. 5) suggests that the model predictions represent the best attainable performance. Thus, the model predictions may not be considered to represent the average pilot's control behavior but the best pilot's performance (which is in agreement with the model assumptions).

A further study (Ref. 5) showed that differences in control behavior between subjects can be "explained" in terms of two model parameters: the indifference threshold ratio (TH) and the overall level of attention (P_o). Both parameters reflect personality traits, such as control strategy/ motivation.

In summary, the experimental results, obtained so far, indicate that the model provides a meaningful representation of pilot workload and dynamics in complex control tasks.

PHYSIOLOGICAL MEASURES

General

Despite a large number of unsolved problems connected with the reliability and validity of physiological measures (Refs. 8, 9, 10), there are still many indications that several physiological parameters contain valuable information regarding the influence of the task on the human operator. As is stated by Roscoe (Ref. 8) :"... there is evidence to support the validity of using heart rate to assess levels of workload in flight but it is necessary to be aware of its limitations". The following discusses some results of NLR research on the adequateness of several physiological measures to be used as indicators of the human operator's activation level.

Experimental results

The physiological measures which have been used in NLR studies on workload include :

- Heart rate parameters (in beats per minute) : mean (M), root mean square (RMS), standard deviation (SD), root mean square of successive differences (RMSSD), and average band power (log ABP).
- Skin resistance parameters : resistance response (SSR) and resistance/ conductance level (SSR/SCL).
- Respiration frequency (inhalations per minute).

The measured values of these parameters have been compared extensively with the outcomes of subjective effort ratings, performance indices, control activity indicators and model workload predictions. A representative example is given in terms of the results which were obtained for the helicopter flying tasks mentioned in the foregoing (see also Ref. 11). The tables 2 and 3 show the intercorrelations between the measured values of the variables for the (relatively easy) navigation tasks and the (relatively difficult) hover task, respectively. These tables reflect a very good correspondance between subjective effort ratings and the physiological parameters of heart rate and respiration. Especially respiration frequency seems to be a remarkable sensitive parameter. With respect to the heart rate measures, the tables indicate that the group heart rate level and heart rate variability (RMSSD), is well correlated with the results from subjective ratings. (Both heart rate measures can be obtained relatively easy). The skin resistance (conductance) measures seem less adequate to be used as workload level indicators. Furthermore, note that the heart rate and respiration measures are also well correlated with the performance index J (mentioned in the foregoing) and with the control activity indicators (longitudinal and lateral cyclic, pedals and collective pitch, respectively). Note, besides, that the correlation between performance index J and the subjective ratings are good for the relatively difficult hover task only (reflecting near optimal control behavior for this task).

Physiological measurements have also been performed in a variety of monitoring tasks under laboratory conditions (Ref. 12), thus imposing only mental workload upon the subjects (general aviation pilots).
The results clearly indicated that all heart rate measures (M, RMS, SD, RMSSD) were only sensitive to the number of displays to be monitored.
The skin resistance response measures (M, RMS, SD, RMSSD) and, to a less degree, the skin resistance level measures appeared to be only sensitive to the aspect of failure uncertainty (presence/absence or prior knowledge about failure type), thus reflecting emotional aspects of the task.
So, there is a good support for using heart rate measures (especially heart rate level and heart rate variability), as a useful tool to study pilot workload in control as well as in monitoring tasks (e.g. the automatic approach). The usefulness of skin resistance measures seems to be restricted to (emotional) aspects of task uncertainty.

CONCLUDING REMARKS

Experimental results obtained under laboratory and in-flight conditions support the usefulness of the optimal control model structure to describe pilot control behavior.

Furthermore, the control effort model predictions have been supported by subjective ratings and physiological measures. Although additional experimental support for the model is desirable, it seems to provide a meaningful representation of pilot workload in complex control tasks.

The physiological measures of heart rate and respiration seem to correlate well with subjective workload ratings, overall performance, measures, control activity indicators and model results. The skin resistance measures tend to be sensitive to uncertainty aspects of the task, more than to performance aspects.

REFERENCES

1	Baron, S et al	Application of optimal control theory to the prediction of human performance in a complex task. AFFDL-TR-69-81, March 1970.
2	Kleinman, D.L. Baron, S.	Manned vehicle system analysis by means of modern control theory. NASA CR-1753, June 1971.
3	Wewerinke, P.H.	Effort involved in single- and two-axis manual control systems. NLR TR 75060 U, November 1974.
4	Kahneman, D.	Attention and effort. Prentice - Hall Inc. 1973.
5	Wewerinke, P.H.	Performance and workload analysis of in-flight helicopter missions. NLR MP 77013 -U, April 1977.
6	Levison, W.H. et al	Studies of multivariable manual control systems. A model for task interference. NASA CR-1746, May 1972.

7 McDonnell, J.D.
Pilot rating techniques for the estimation and evaluation of handling qualities.
AFFDL TR 68-76, December 1968.

8 Various authors
Assessing pilot workload.
AGARDograph No. 233, February 1978.

9 Heslegrave, R.J. et al
Measuring baseline-treatment differences in heart rate variability: variance versus successive difference mean square and beats per minute versus interbeat intervals.
Psychophysiology, Vol. 16, No. 2, 1979.

10 Wastell, D.G.
Measuring heart rate variability: comments on the successive difference mean square statistic.
Psychophysiology, Vol. 18, No. 1, 1981.

11 Smit, J.
Wewerinke, P.H.
An analysis of helicopter pilot control behavior and workload during instrument flying tasks.
NLR MP 78003 U, February 1978.

12 van de Graaff, R.C.
An experimental analysis of human monitoring behavior in multivariable failure detection tasks.
NLR TR 81063, July 1981.

SUBJECT	TASK	COMPUTED EFFORT (dB)	SUBJECTIVE RATING		
			AVERAGE	EFFORT	DEMAND
A	HOVER	17.0	5.3	5.0	5.5
	NAV-L	15.9	4.9	4.5	5.4
	NAV-H	15.9	4.6	4.3	4.8
B	HOVER	16.5	7.5	7.5	7.4
	NAV-L	15.1	6.4	6.4	6.4
	NAV-H	14.9	6.7	6.7	6.7
Average of 4 subjects	NAV-L	15.0	5.3	5.2	5.5
	NAV-H	15.1	5.4	5.2	5.7

TABLE 1: Comparison of control effort model results and subjective ratings

(Adapted from reference 5)

TABLE 2

Intercorrelation matrix for the high and low navigation trials (42 runs)

	Physiological variables						Performance index (J)	Subjective ratings			Control activity			
	HR	RMSSD	log ABP	Resp. freq.	SCL	SRR		Safety pilot rating	Effort rating	Demand rating	σ_{δ_e}	σ_{δ_a}	σ_{δ_r}	σ_{CP}
Heart rate	-	-.42*	-.56*	.67*	-.05	-.30	.38*	.57*	.14	-.01	-.58*	-.14	-.66*	-.69*
RMSSD	-.42*	-	.51*	.01	-.40*	.43*	.34*	-.16	.48*	.43*	.16	-.04	.56*	.53*
log ABP	-.56*	.51*	-	-.38*	.22	-.04	-.19	-.38*	.19	.21	.27	.09	.43*	.41*
Respiration frequency	.67*	.01	-.38*	-	.06	-.19	.39*	.50*	.63*	.37*	-.24	-.10	-.18	-.26
SCL	-.05	-.40*	.22	.06	-	-.51*	-.33*	-.20	.14	.15	.09	-.00	.08	.09
SRR	-.30	.43*	-.04	-.19	-.51*	-	.29	-.02	-.13	-.06	.08	-.06	.18	.19
Performance index (J)	.38*	.34*	-.19	.39*	-.33*	.29	-	.33*	.29	.28	-.25	-.11	-.01	-.02
Safety pilot rating	.57*	-.16	-.38*	.50*	-.20	-.02	.33*	-	.12	.03	-.12	-.24	-.44*	-.44*
Effort rating	.14	.48*	.19	.63*	.14	.13	.29	.12	-	.76*	-.11	-.10	.20	.10
Demand rating	-.01	.43*	.21	.37*	.15	-.06	.28	.03	.76*	-	-.09	-.31*	.15	.10
σ_{δ_e}	-.58*	.16	.27	-.24	.09	.08	-.25	-.12	-.11	-.09	-	.09	.64*	.61*
σ_{δ_a}	-.14	-.04	.09	-.10	-.00	-.06	-.11	-.24	-.10	-.31*	.09	-	.12	.18
σ_{δ_r}	-.66*	.56*	.43*	-.18	.08	.18	-.01	-.44*	.20	.15	.64*	.12	-	.88*
σ_{CP}	-.69*	.53*	.41*	-.26	.09	.19	-.02	-.44*	.10	.10	.61*	.18	.88*	-

* Values of r exceeding .30 are significant ($p < .05$)

(Adapted from reference 11)

TABLE 3

Intercorrelation matrix for the hover trials (33 runs)

	Physiological variables						Performance index (J)	Subjective ratings			Control activity			
	HR	RMSSD	log ABP	Resp. freq.	SCL	SRR		Safety pilot rating	Effort rating	Demand rating	σ_{δ_e}	σ_{δ_a}	σ_{δ_r}	σ_{CP}
Heart rate	-	-.51*	-.41*	.47*	.01	-.32	.44*	.67*	.47*	.44*	.19	.39*	.28	-.09
RMSSD	-.51*	-	.75*	-.45*	-.25	.28	-.46*	-.47*	-.29	-.46*	-.20	-.44*	-.08	.12
log ABP	-.41*	.75*	-	-.73*	.10	.08	-.59*	-.70*	-.61*	-.73*	-.19	-.45*	.11	.31
Respiration frequency	.47*	-.45*	-.73*	-	.16	-.40*	.43*	.70*	.77*	.58*	.03	.18	-.13	-.44*
SCL	.01	-.25	.10	.16	-	-.66*	-.33	-.30	.07	-.16	-.19	-.24	-.01	-.18
SRR	-.32	.28	.08	-.40*	-.66*	-	.02	-.05	-.18	.03	.18	.18	-.03	.18
Performance index (J)	.44*	-.46*	-.59*	.43*	-.33	.02	-	.65*	.41*	.63*	.26	.48*	-.03	-.18
Safety pilot rating	.67*	-.47*	-.70*	.70*	-.30	-.05	.65*	-	.67*	.67*	.23	.42*	.18	-.13
Effort rating	.47*	-.29	-.61*	.77*	.07	-.18	.41*	.67*	-	.80*	.19	.38*	.10	-.34
Demand rating	.44*	-.46*	-.73*	.58*	-.16	.03	.63*	.67*	.80*	-	.38*	.58*	.09	-.22
σ_{δ_e}	.19	-.20	-.19	.03	-.19	.18	.26	.23	.19	.38*	-	.87*	.53*	.34
σ_{δ_a}	.39*	-.44*	-.45*	.18	-.24	.18	.48*	.42*	.38*	.58*	.87*	-	.46*	.22
σ_{δ_r}	.28	-.08	.11	-.13	-.01	-.03	-.03	.18	.10	.09	.53*	.46*	-	.65*
σ_{CP}	-.09	.12	.31	-.44*	-.18	.18	-.18	-.13	-.34	-.22	.34	.22	.65*	-

* Values of r exceeding .34 are significant ($p < .05$)

(Adapted from reference 11)

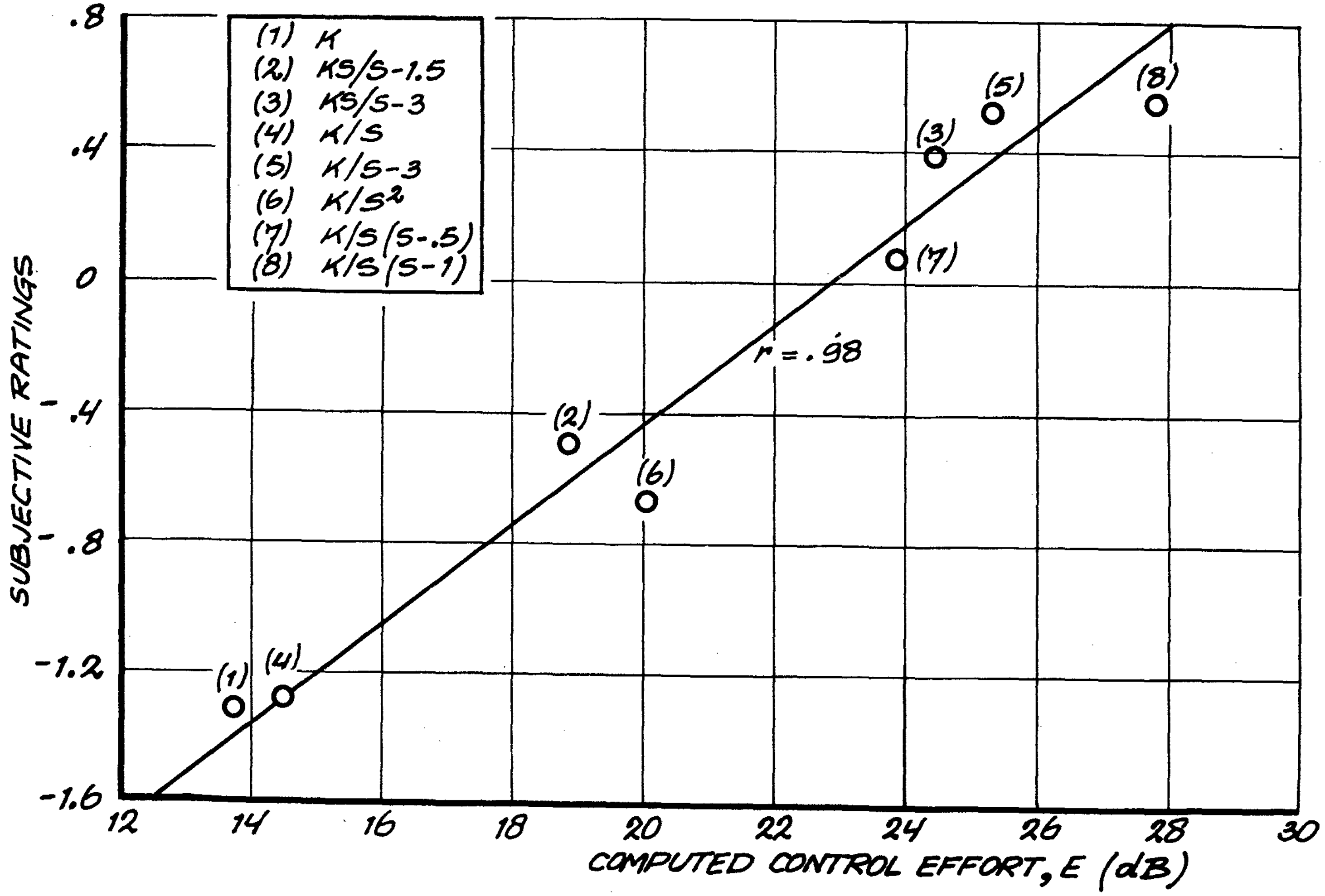


Fig. 1 A comparison of computed control effort and subjective ratings for the single-axis configurations

(Adapted from reference 3)

RESTRAINT CONSIDERATIONS IN DYNAMIC ENVIRONMENTS

Dana B. Rogers
John W. Frazier

Air Force Aerospace Medical Research Laboratory
Aerospace Medical Division, Air Force Systems Command
Wright-Patterson Air Force Base, Ohio

ADDRESS FOR CORRESPONDENCE: AFARML/BBS
Wright-Patterson AFB, OH 45433

ABSTRACT

The pilot's workload and ability to control the aircraft are often influenced by the restraint system. While the need for restraint systems is obvious, criteria for good restraints are often in conflict. For example, it is desired to provide the pilot with a system which provides multi-directional support while also allowing unencumbered mobility. Other considerations include comfort, ease of use, cockpit compatibility, visibility and other factors.

The Dynamic Environment Simulator (DES), a three-axes human centrifuge, has been used to simulate a wide range of acceleration environments. An overview of Gx, Gy, and Gz restraint techniques, problems, and tolerance will be presented. Specific simulations (supported with movies when available) of the F-15 and KFIR-C2 flat spin environments, B-1 escape, and AFTI/F-16 lateral maneuvering will be discussed. The pilot's ability to perform control inputs or initiate ejection sequences will be discussed. Active restraints in the form of pneumatic supports or variable tension straps may be techniques for providing improved restraint for unconventional flight environment.

THE EFFECTS OF SUSTAINED ACCELERATION, AIRFRAME BUFFET, AND AIRCRAFT FLYING QUALITIES ON TRACKING PERFORMANCE

A. G. Piranian*
U.S. Naval Air Development Center
Warminster, Pa.

Abstract

To determine the relative importance of some of the factors affecting the air-to-air tracking performance of a pilot airframe system, U.S. Navy test pilots performed tracking tasks in a centrifuge-mounted F-4 cockpit plus visual display. Sustained accelerations up to 5.0g, buffet intensities of up to ±0.5g, and lateral-directional flying qualities were varied independently, tracking performance being measured in terms of miss distance, percentage time within a fixed radius of the target, and pilot opinion ratings. Results show a dominant influence of flying qualities, significant influence of sustained acceleration, and negligible influence of airframe buffet.

Introduction

At the high speeds and angles-of-attack normally attained in air combat maneuvering flight, high sustained acceleration, airframe buffet, and aircraft flying qualities combine to increase significantly the difficulty of the fighter pilot's air-to-air tracking task. The literature concerning the relative magnitudes of these influences is sparse and inconclusive. Existing data on human tolerance and performance in a buffeted environment are limited to 1-g flight and are inconsistent due to nonuniformity in restraint systems, seats, clothing equipment, control task and scoring methods.(1) Existing data concerning the influence of various levels of flying qualities on tracking performance are, for the most part, qualitative and do not separate flying qualities effects from buffet and g effects. All results taken together would seem to indicate that flying qualities influence tracking ability considerably, that buffet has a lesser, though appreciable influence, and that sustained medium -to-high g has an insignificant effect. The present study was undertaken to determine, quantitatively, just how much of an influence each of these factors has.(2)

Analysis

Since these factors are difficult to vary independently in flight and since independent variation of them during the same tracking task was necessary for this study, it was decided to perform the investigation in a ground-based simulator. The Naval Air Development Center human centrifuge flight simulator was particularly well suited to the requirements of the study, combining the capabilities of a fixed-base simulator with that of imposing a wide range of sustained acceleration levels.

Simulator

The centrifuge consists of a 10 ft diameter spherical gondola supported in a two gimbal system on an arm of 50 ft radius. The arm rotates counter clockwise, with the gimbals permitting the gondola (and, hence, the cockpit within the gondola) to both pitch and roll. The gimbals permit orientation of the cockpit so that the centrifugal force vector, due to rotation of the arm, is coincident with the resultant of the forces experienced in actual flight. The cockpit was mounted facing the direction of arm rotation, resulting in the rotation of the simulator pilots towards the hub of the centrifuge to generate positive normal load factors. The forward end of the cockpit was supported on a pivot mounted to the gondola, while the rear end was supported by lateral and vertical hydraulic shakers as shown in Figure 1.

The cockpit was an F-4B Phantom cockpit and had been removed from a strike damaged aircraft. The interior duplicated that of an F-4B except for the instrument panel which had been modified slightly for a previous stall/spin simulation program. The flight controls were set up to duplicate the force/control surface deflection characteristics of a current high performance fighter, and stability augmentation in all three axes could be switched on or off.

A visual display was provided in the area bounded by the front windscreen. The clear canopy was replaced by a dark, opaque one to restrict the pilot's visual cues to within the cockpit (i.e., instruments and visual display). An all-attitude, full color display of earth, horizon, and sky was provided through a lens/Attitude Director Indicator (ADI) system.

---

*Aerospace Engineer, Air Vehicle Technology Department

The ADI was painted to provide a blue sky with clouds, a distinct horizon line, and ground with clutter to simulate populated areas. The zoom lens transmitted the ADI display via a mirror to a projection screen in the cockpit windscreen, as seen in Figure 2. The lens was used to create a realistic sense of altitude and altitude changes.

Target and gunsight images were superimposed on the terrain display in a beam splitter. A TV camera projected these images from a cathode ray tube onto the beam splitter. The target was represented by an inverted "T", and the gunsight was a fixed reticle with horizontal and vertical reference marks, as shown in Figure 3. The entire display was viewed through a fresnel lens to provide focus at infinity.

To make the cockpit environment as realistic as possible, the simulator pilots were asked to wear full flight gear: flight suit, boots, helmet, oxygen mask, torso harness, g-suit, and full life support equipment. The valve which inflates the pilot's g-suit in proportion to normal acceleration was functional and used during all dynamic runs. Additionally, pressure breathing air was supplied to the pilot's oxygen mask from diving tanks. He was able to control cockpit temperature through air conditioning controls on the right console.

Three Electronic Associates Incorporated 231-R analog computers were used to run the simulation. Aircraft equations of motion, control system dynamics and centrifuge coordinate transformations were all programmed on the analog computers.

Task

To provide the simulator pilots with as realistic a task as possible while simultaneously minimizing additional influences on pilot performance due to the task itself, the target maneuver selected was a bank into a level, constant load factor turn, a short time in the steady state turn, and a bank back to straight, level flight. In terms of time, the roll to the prescribed bank angle took 5 sec, 15 sec were spent in the steady state turn, and rollout back to straight flight took another 5 sec.

Experience has shown that air combat maneuvering is centered in the transonic speed regime. The flight condition simulated was, therefore, Mach .8 at an altitude of 10,000 ft, with the simulated aircraft 1500 ft from the target aircraft and at his 6 o'clock position initially. The target represented an aircraft with a 50 ft wingspan.

The ranges of buffet and sustained normal loads were selected to encompass extremes encountered in flight. Thus buffet intensity was varied from 0 to ±.5g (severe buffet), where buffet intensity levels are in terms of peak to peak values, and sustained normal loads were varied from 1.3g to 5.0g. 1.3g was selected as a lower limit rather than 1.0g to permit continuous rotation of the centrifuge, while 5.0g was found to be an upper physiological limit of continuous load tolerable for 25 sec without grayout. Discrete values used within the respective ranges were ±.1g and ±.25g buffet, and 2.5g sustained load.

Each task then consisted of a random right or left turn of the target to the bank angle required to obtain 1.3, 2.5 or 5.0g in level flight, with buffet and sustained load on the cockpit building in the bank, remaining constant for 15 sec in the steady state turn, and decreasing again in the final 5 sec while the target rolled back to straight, level flight. It should be noted that regardless of pilot stick and or throttle inputs, the target remained at a constant 1500 ft range.

Performance Criteria

Tracking performance was measured in terms of projected miss distance from the target, percentage time within 10 mils of the target, and pilot opinion ratings, using the Cooper-Harper scale. During the simulation, brush recorders printed out time histories of the aircraft motion variables as well as weapon/target lateral offset distance, vertical offset distance, and the root mean square offset integrated over 5 sec time segments. Pilot opinion ratings were given at the conclusion of each 25 sec task, with ratings reflecting the pilot's assessment of the run in terms of his ability to keep the pipper on or return the pipper to the target. All pilot comments were recorded on tape during the simulation.

Pilots

Eleven test pilots from the Naval Air Test Center, Patuxent River, Maryland performed the simulation. All were experienced fighter or attack pilots, having flown one or more of the various F-4, F-8, A-4 and/or A-7 aircraft. During the simulation, pilots were observed via a low light TV camera mounted in the right side of the cockpit, voice communication was maintained through the "hot" microphone in the pilots' oxygen masks and electrocardiogram signals were continuously observed by the medical monitor.

Buffet Phase

The simulation was divided into two phases. In the first phase the effects of buffet and sustained normal acceleration on tracking performance were measured. In the second phase, the effects of flying qualities were determined. To isolate buffet and sustained acceleration effects, flying qualities remained invariant in the first phase (stability derivatives in the six degree-of-freedom equations of motion of a rigid body were held constant) and were the low angle-of-attack flying qualities of a current high performance fighter aircraft.

Flight tests of current fighters with cockpit-mounted accelerometers show that lateral and vertical component buffet peaks occur at very nearly the same frequencies, with lateral intensity being equal to or less than the corresponding vertical intensity. The primary resonance frequency excited in each aircraft was that corresponding to the first wing bending mode of that aircraft. Power spectral density plots of the component accelerations of the A-6, F-104, F-8A, F-5A, F-106A, and F-4E shown as functions of frequency indicate that resonance peaks occur at frequencies between 5 cps and 18 cps. (3)(4)(5) To limit the simulation to a reasonable length, one representative buffet frequency was selected. Since several of the above aircraft exhibited peaks in the 10 cps area, that frequency was used throughout the simulation. Equal frequency and intensities were applied vertically and laterally.

To provide an unbiased, consistent data base, combinations of buffet, sustained acceleration, and target turn direction were randomized so that each set of 12 simulator runs encompassed all possible combinations of buffet and sustained acceleration. Six pilots participated in this phase of the simulation. Each pilot was scored for 4 sets of runs, or 48 total runs. Thus 288 scoring runs were made altogether, with 24 runs being made for each desired data point.

Flying Qualities Phase

The literature was researched to determine fighter aircraft flying qualities in the higher angle-of-attack range experienced in air combat. Angles-of-attack up to 25° were considered for the F-4, F-5, F-8, F-104, and F-106. In general it was found that:

Longitudinal flying qualities change very little
Aileron effectiveness in roll decreases
Rolling due to yawing increases
Directional stability changes from positive to negative
Aileron yaw changes from proverse to adverse

The only fighter aircraft showing a noticeable variation in longitudinal flying qualities at high angles-of-attack was the F-104. A reduction in pitch damping was found to occur as normal acceleration was increased. This is attributable to the unconventional "T" tail configuration of that aircraft, the horizontal tail being immersed in the wing-fuselage wake at higher angles-of-attack.

It was concluded that the dominant influence on aircraft maneuverability at high angles-of-attack was the deterioration in lateral-directional stability and control. Serious problems noted for several of the above-mentioned aircraft are wing rock and nose slice, where wing rock is defined as an undamped or divergent dutch roll with very large roll excursions, and nose slice is a yawing oscillation caused by low or negative directional stability. The F-4, in particular, has been shown to exhibit a progression of dutch roll from a lightly damped rolling oscillation (up to ± 40° roll excursions) to an unstable yawing oscillation at 23° angle-of-attack.(6)(7)

To measure the influence of certain lateral-directional stability characteristics on tracking performance, the simulator's stability augmentation system was turned off, first, the simulator was flown without augmentation to obtain reference tracking scores, and, finally, high angle-of-attack lateral-directional stability derivatives were set in the analog computer. (Note that, again, these derivatives remained constant regardless of angle-of-attack commanded by the pilot.) Due to the significant influence of aileron yaw on tracking performance demonstrated in tracking tests at NASA Flight Research Center and at Calspan Corporation, the effects of proverse, zero, and adverse aileron yaw were also investigated.(8)

Before the simulator could be flown with stability augmentation off, adjustments had to be made to the yaw damping derivative, although it was the actual derivative of a current, high performance fighter (as were all coefficients). Dutch roll damping was so light that the target could not be kept within the visual display by the pilots. Yaw damping was gradually increased until the pilots indicated acceptability, raising the dutch roll damping ratio from .1 to .31.

Stability derivatives changed for the high angle-of-attack case were dihedral effect, roll damping, roll due to yaw rate, yaw due to roll rate (all of whose magnitudes were increased), and directional stability (decreased). Again one of the derivatives required modification. The actual directional stability of the fighter which was simulated decreases with increasing angle-of-attack, passing through zero to negative values above 23° Pilots were unable to track with directional stability at or below zero. The derivative was therefore increased gradually until an acceptable value was found.

In evaluating the effects of aileron yaw, the reference configuration was used, varying only the coefficient of roll due to roll control from positive to zero to negative.

To allow a comparison of results with those of the first phase, sustained normal accelerations of 1.3, 2.5, and 5.0g were used. Two vibration levels were imposed - no buffet and ±.25g. The former allowed assessment of flying qualities influences, alone, on tracking, while the latter provided a reference buffet condition for comparison with the first phase results, allowing assessment of tracking error increment under the combined influences of all variables.

Combinations of variables were randomized so that a set of 6 simulator runs encompassed all combinations without bias. Five pilots participated in this phase of the simulation, with each pilot required to perform 3 sets of runs each in the reference and high angle-of-attack configurations, and two sets of runs at each of the remaining roll control conditions, resulting in 60 runs per pilot.

Results

Buffet Phase

As noted previously, tracking performance was measured in terms of projected miss distance in root mean square (RMS) feet, percentage time within 10 mils of the target, and pilot opinion (Cooper-Harper) ratings. Note that mil is defined here as one foot radial offset per thousand foot range, or since range is 1500 feet, 10 mils represent a 15 foot error. RMS error was integrated over 5 second intervals of the 25 second task. Variables of interest in this phase were the steady state acceleration and buffet levels, or the middle 15 seconds of the task. Shown in figure 4 are the RMS scores of all six pilots plotted versus normal acceleration as a function of buffet level. Since RMS was integrated over 5 second segments each pilot's scores were averaged over the 15 seconds at steady state, and then all pilots' scores were averaged together. Error was recorded only in those instances in which the target remained in the visual display over the prescribed time period. Some runs were not scored due to loss of the target.

Figure 4 indicates that sustained acceleration had a pronounced effect on tracking performance while the effect of buffet was negligible. There is an approximately linear increase in error with respect to normal acceleration of 6 feet RMS per g, and almost no change in error with respect to buffet intensity except at 5g. The scatter at 5g is due to three causes:

1. Smaller data base due to loss of target
2. Variation in pilot "learning curve" or time to adapt to the simulator's lack of stick force per g feel
3. The lack of buffet or minimal buffet at high g felt unrealistic to the pilot, leading to decreasing error with increasing buffet.

At 1.3 and 2.5g the target was virtually never lost, while at 5.0g the target was lost 80% of the time.

Figures 5 and 6 reflect the same conclusions as figure 4. Figure 5 shows percentage time within 10 mils of the target versus sustained normal acceleration as a function of buffet level. While percentage time is scattered over a 10% range from no buffet to very severe buffet at a particular acceleration level, it decreases approximately linearly with increasing sustained acceleration at a rate of 20% per g.

Figure 6 shows a barely perceptible increase in pilot opinion rating (POR) with increasing buffet intensity, and an approximately linear increase of POR with sustained normal acceleration at the rate of .75 per g. Up to 2.5g the aircraft was found to be satisfactory, thence going to moderately objectionable at 5.0g. Particularly noteworthy is the correspondence of POR to RMS error. Figure 7 indicates a linear correspondence between the two.

Pilot comments regarding this phase of the simulation were as follows:

> Simulator handling qualities were realistic except for too great a lateral stick sensitivity
>
> Lateral stick sensitivity led to mild PIO's
>
> Tracking field of view was too small
>
> Lack of minimal buffet at high sustained g felt unrealistic and abnormal
>
> Buffet feel corresponded very well to actual aircraft buffet.

Prior to the simulation, pilots suspected buffet as being unimportant in its direct effect on tracking performance; insofar as it is a cue to imminent stall/

departure it does, however, divert pilot attention from the target to his own aircraft.

Pilots displayed a high degree of confidence in the results of the simulation - i.e. that normal acceleration had a more pronounced effect on pilot tracking ability than did buffet

Absolute quantitative conclusions must be qualified by noting that buffet frequency (10 cps) was much higher than aircraft response modes; at lower frequencies buffet might have more of an effect on performance.

Regarding the complaints of too much lateral stick sensitivity, it should be noted that this is a deficiency characteristically noted on fixed base simulators and is attributable to lack of roll rate cues to stick inputs.

Flying Qualities Phase

Flying qualities affected the tracking solution primarily in the initial acquisition of the target. Once the target was in the steady state turn, the tracking task became essentially non maneuvering. To properly assess flying qualities' influences, therefore, the initial 5 seconds of the task were scored. A comparison was obtained by plotting the tracking scores of the six pilots from the previous phase for the initial 5 seconds of the task, as shown in figure 8. Since buffet and acceleration were building to steady state levels in that segment, somewhat better performance was achieved than at steady state.

From the beginning of this phase of the simulation it was evident that flying qualities would be the factor of paramount importance in tracking accurately. The inability of the pilots to track with the actual yaw damping and directional stability of the simulated aircraft clearly demonstrated this.

Interestingly, the minimum dutch roll damping requirements of Mil-F-8785B were inadequate to permit satisfactory tracking. With stability augmentation off, both the spiral and roll modes satisfied level 1 requirements for a class IV aircraft in flight phase category A.(9) Dutch roll damping ratio, 0.1, met the level 2 requirements. The pilots could not successfully track the target until the damping ratio had been raised to 0.3, a value well above level 1 requirements, .19. McDonnell Douglas had also found that air-to-air tracking was impossible at a damping ratio of .1, difficult but possible at .3, and good at .5.(10)

Time histories of the control inputs indicated that the pilots attempted to track the target using elevator and aileron control alone, until established in the steady state turn. Therefore the bank angle to aileron deflection transfer function was considered for each configuration.

For the reference and high angle-of-attack configurations the dutch roll loop closures are destabilizing. For the latter configuration it is seen that the loop closure almost goes to neutral stability, making it obvious why directional stability had to be increased over its original value. At values less than that which was used, the loop closure enters the unstable half of the plane, accounting for pilot inability to track at those values.

It was expected that the high angle-of-attack configuration would yield poorer performance than the reference aircraft because of the greater difficulty in controlling bank angle with marginal closed loop dutch roll stability. Figure 9 compares the tracking scores and, surprisingly, shows that performance was not significantly different. At 5.0g the high angle-of-attack flying qualities led to an approximately 10 ft greater error, while at the lower g performance was virtually the same. However, pilot workload was higher.

Consideration of the dutch roll loop closure in the bank angle to aileron deflection transfer functions for the various yaw due to aileron control configurations reveals that bank angle control is best in the case of zero aileron yaw, followed by adverse yaw and then proverse yaw.

Figure 10 compares the respective tracking scores, showing similar performances with adverse and proverse yaw. Clearly the neutral configuration is superior to the latter two, tracking performance improving considerably - i.e. scores are almost 50% better.

Performance in terms of percentage time within 10 mils of the target shows the same trends (figure 11). The reference and high angle-of-attack conditions produce nearly identical results, proverse yaw is slightly better than adverse yaw, and zero aileron yaw is seen to be a greatly enhancing quality.

Pilot opinion ratings, shown in figure 12, are in slight disagreement with the above results, in that ratings are identical for the reference (proverse yaw) condition and the zero yaw condition. Since performance is better at the latter condition, the ratings may be due to either pilot fatigue (since the zero yaw condition was the last set of runs flown by the pilots) or greater sensitivity in pilot rating as the task became easier. It is easily seen that the high angle-of-attack condition required greater compensation by the pilot than the reference condition, although both were flown equally well. And the adverse yaw condition, although flown almost as well as the proverse yaw condition, required maximum compensation by the pilot to the point where he considered it a major deficiency requiring improvement at high g.

Comparing the reference condition with that of the augmented, phase I aircraft, figures 13-15 show the greater tracking accuracy with the latter aircraft, and the increased pilot compensation required to attain even the reduced accuracy. With stability augmentation off, tracking error increased by more than 15 feet (10 mils) or approximately 100%, time within 10 mils of the target decreased by approximately 33% and POR's were 1 rating worse. With the modified dutch roll damping ratio, zero aileron yaw leads to performance very close to that with stability augmentation on.

Sustained acceleration effects on tracking are the same for the stability augmentation on and stability augmentation off configurations. Figures 13 and 15 show nearly the same slopes for both configurations. Buffet, again, has no significant effect on performance.

The paramount importance of flying qualities in tracking is evident. Not only was tracking impossible at the simulated aircraft's actual yaw damping level and directional stability, but even with these values raised to acceptable levels, tracking was poor. In terms of RMS error, decreasing dutch roll damping ratio from .8 (stability augmentation on) to .3 degraded tracking precision as much as a 4 g increase in sustained acceleration.

Pilot commentary concerning this phase of the simulation was as follows:

> Tracking field of view was too small
>
> Buffet would actually have more of an effect because of its indication of imminent stall departure
>
> Buffet felt realistic
>
> Lateral PIO was caused by low dutch roll damping ratio
>
> For equal proverse and adverse aileron yaw, proverse is preferable
>
> Zero aileron yaw led to the most satisfactory configuration

Results of this phase of the simulation are in good agreement with those of references (8) and (10). Reference (10) shows tracking ability to degrade with decreasing dutch roll damping ratio, tracking being impossible at a value of .1. Tracking tests at NASA FRC showed that wing rock caused tracking errors of almost 30 mils and POR's of 10, performance significantly worse than that due to buffet or sustained accelerations in their tests.

Considering the effects of yaw due to aileron control, NASA FRC showed optimum tracking precision slightly to the proverse side of zero, with pilot comments reflecting a preference for proverse yaw. Reference (8) also indicates a preference by pilots for small proverse yaw, although optimum air-to-ground weapons delivery occurs on the adverse side of zero. These results correspond with those of the present study, where optimum precision occurs at approximately zero yaw due to aileron and where pilots prefer proverse to adverse yaw.

Conclusions

1. Buffet intensities of up to ±.5g at a frequency of 10 cps have a negligible influence on tracking precision.

2. Sustained, high normal accelerations of up to 5.0g appreciably degrade tracking precision.

3. Flying qualities have the most significant influence on tracking precision.

4. Dutch roll frequency and damping ratio have a serious influence on tracking precision; tracking error increases as either quantity decreases, to the point where tracking becomes impossible at values normally encountered at high angles-of-attack.

5. The minimum dutch roll damping requirements of MIL-F-8785B appear to be inadequate to accomplish the air-to-air tracking mission.

6. For equal amounts of adverse or proverse yaw due to lateral control input, tracking precision is approximately the same, although pilots prefer proverse yaw.

7. Providing a pilot with roll control that eliminates yaw due to lateral control input significantly enhances air-to-air tracking capability.

8. Tracking precision can best be improved by improving basic aircraft stability and control at high angles-of-attack.

References

(1) Eney, J. A., "Buffet in Air Combat Maneuvering: Problem Definition," Naval Air Development Center Report NADC-AM-7040, 23 September 1970

(2) Piranian, A. G., "The Influences of Buffet, Sustained Normal Accelerations, and Basic Aircraft Flying Qualities on Tracking Performance in Air Combat Maneuvering," Naval Air Development Center Report NADC-73257-30

(3) Titiriga, A., Jr. "F-5A Transonic Buffet Flight Test," Northrop Corporation Report AFFDL-TR-69-110, December 1969

(4) Emerson, E. R., "F-106A Transonic Buffet Flight Test," Convair Report AFFDL-TR-70-87, June 1970

(5) Cohen, M., "Buffet Characteristics of the Model F-4 Airplane in the Transonic Flight Regime," McDonnell Aircraft Company Report AFFDL-TR-70-56, April 1970

(6) Rutan, E. L., and Gentry, J. R., Maj., USAF, "Category II Stability and Control Evaluation of the F-4E Aircraft," Air Force Flight Test Center Report FTC-TR-69-14, April 1969

(7) Chambers, J. R., and Anglin, E. L., "Analysis of Lateral-Directional Stability Characteristics of a Twin-Jet Fighter Airplane at High Angles of Attack," NASA Technical Note TND-5361, August 1969

(8) Hall, G. W. Weingarten, N. C., and Lockenour, J. L., "An In-Flight Investigation of the Influence of Flying Qualities on Precision Weapons Delivery," AIAA Paper No. 73-783, Aircraft Design, Flight Test, and Operations Meeting, 6-8 August 1973

(9) Anon, Military Specification, "Flying Qualities of Piloted Airplanes," MIL-F-8785B (ASG), 7 August 1969

(10) Luedde, W. J., Mathews, R. H., and Gruner, L. D., "The Effect of Gunline Orientation and Flight Control Characteristics on Air to Air Tracking," McDonnell Douglas paper presented at AIAA Fighter Aircraft Conference, 5-7 March 1970

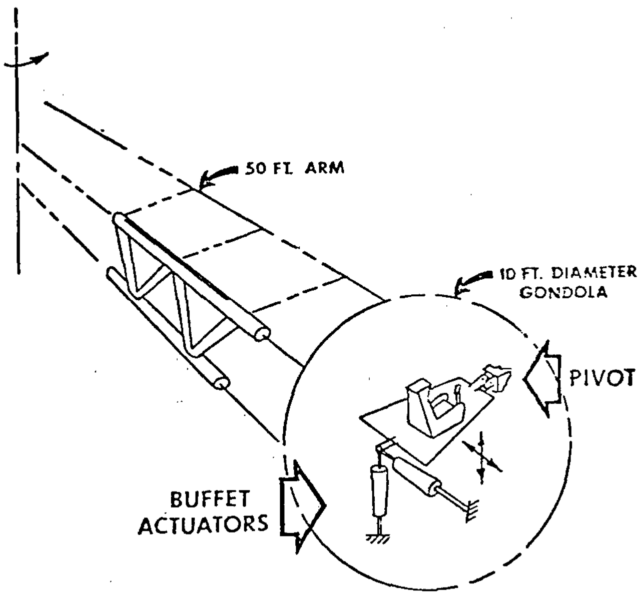


Figure 1. Shaker System on Centrifuge

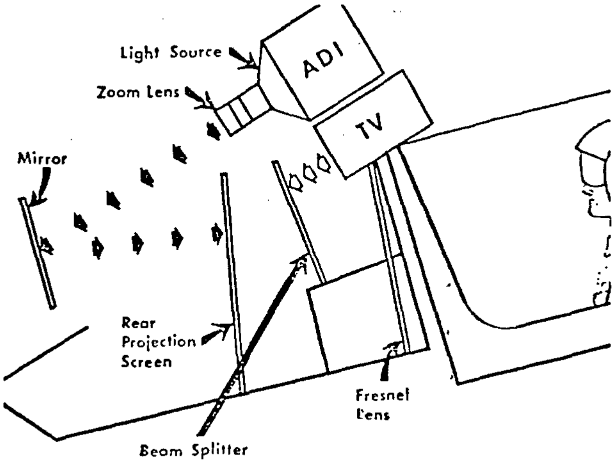


Figure 2. Combination of Terrain and Target Visual Display

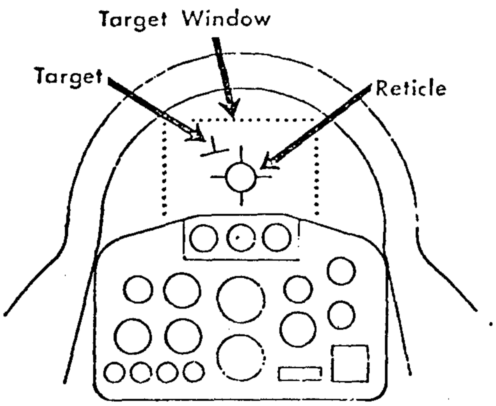


Figure 3. Tracking Display Arrangement

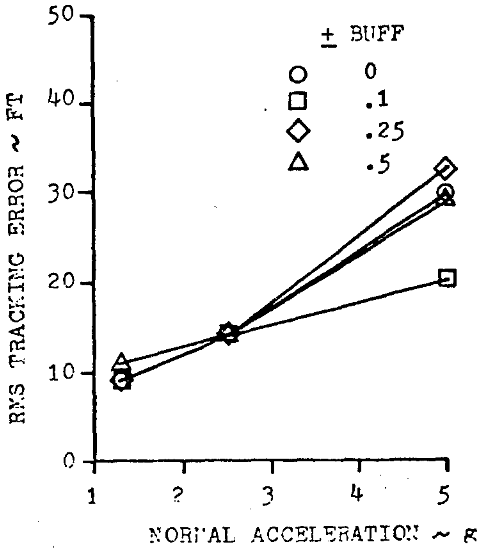


Figure 4. Normal Acceleration Effects

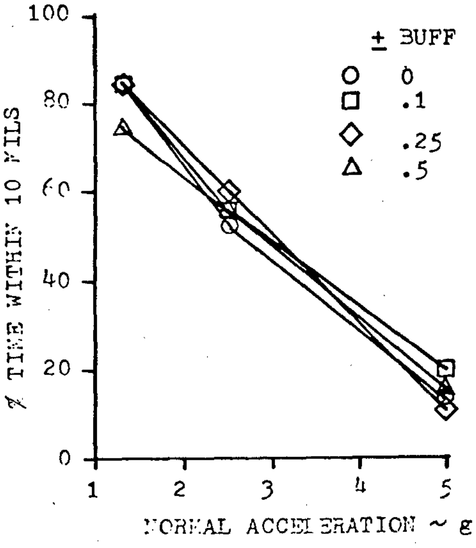


Figure 5. Tracking Performance in % Time

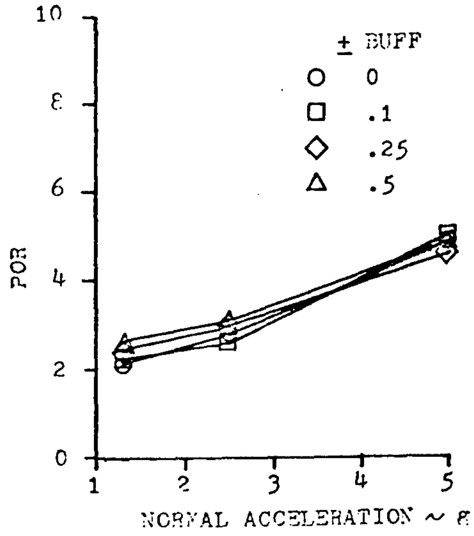


Figure 6. Pilot Opinion Ratings - Normal Acceleration

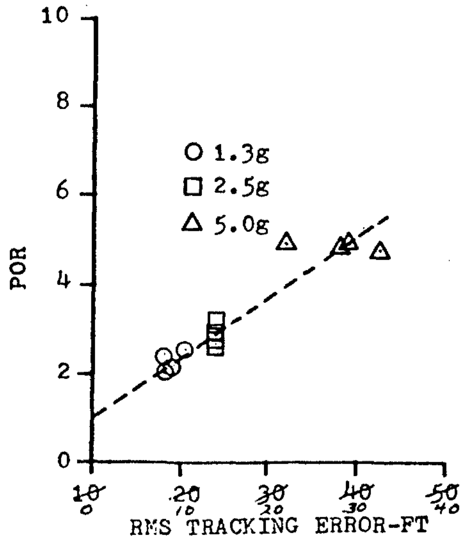


Figure 7. Correspondence of POR and Tracking Error

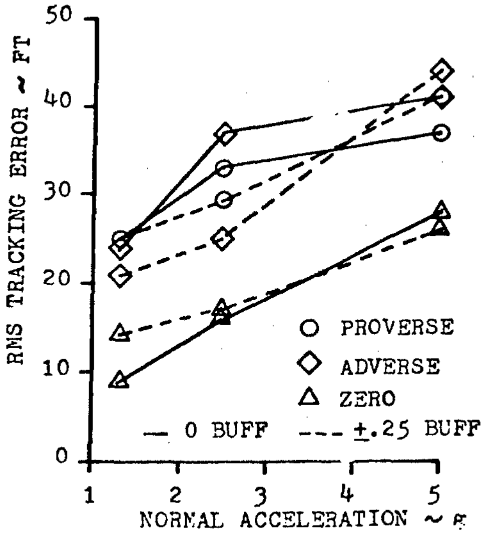


Figure 10. TRACKING SCORES FOR VARIOUS AILERON YAW CONDITIONS

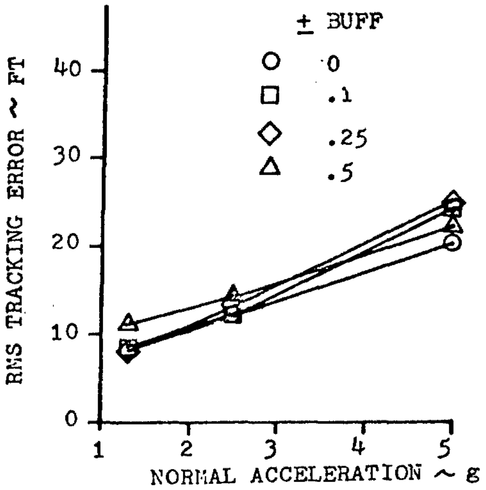


Figure 8. Normal Acceleration Effects-Initial Segment

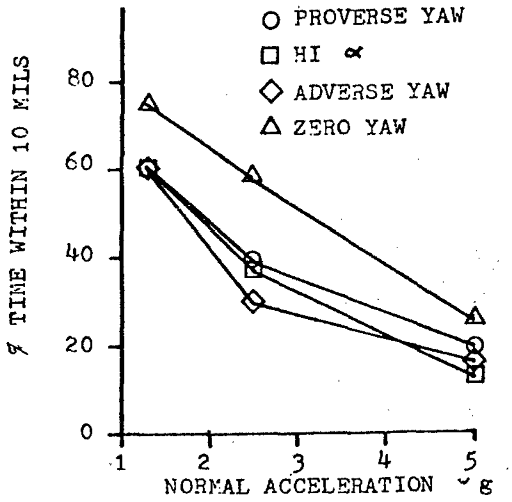


Figure 11. Tracking Performance in % Time

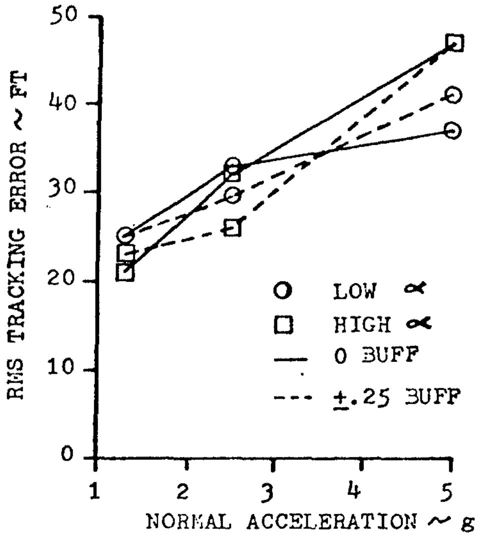


Figure 9. Lo α and Hi α Tracking Scores

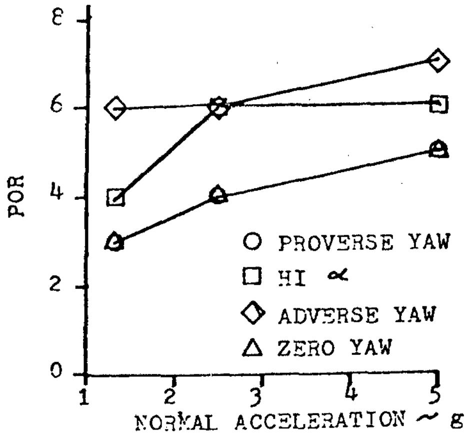


Figure 12. Pilot Opinion Ratings

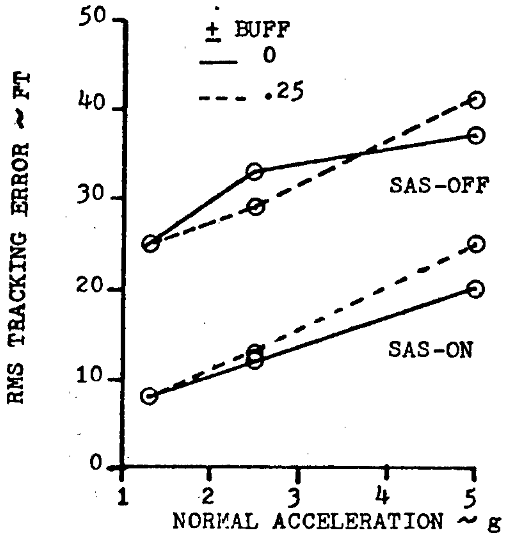


Figure 13. Effects of SAS On Tracking Performance

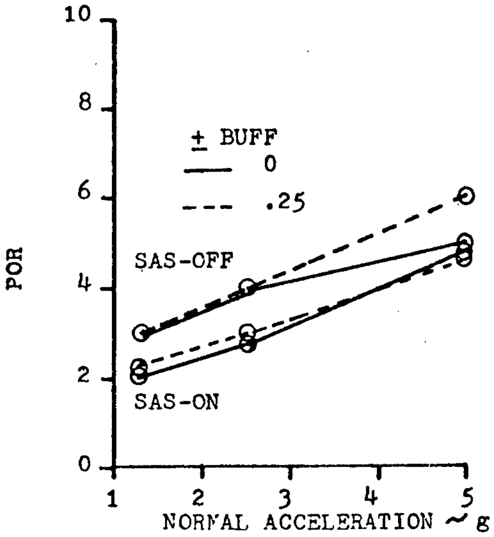


Figure 15. Effects of SAS - POR

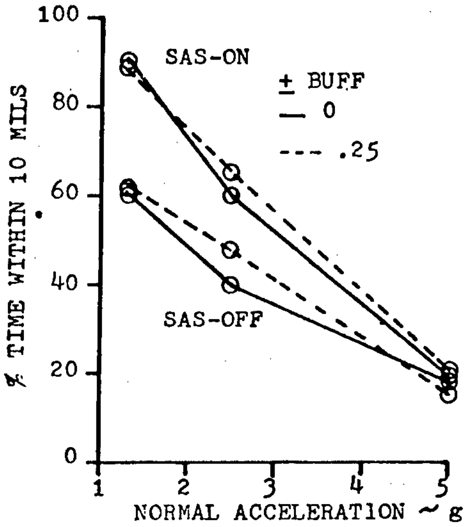


Figure 14. Effects of SAS - % Time

REFLECTIONS ON THE EFFECTS OF VEHICLE DYNAMICS AND TASK DIFFICULTY ON COOPER-HARPER PILOT OPINION RATINGS, TASK PERFORMANCE, AND PILOT WORKLOAD

by

Robert B. Crombie, Captain, USAF
Flight Dynamics Laboratory

This paper is intended to set forth generalized principles on a broad subject which lies at the heart of both flying qualities and human factors research, development, test, and evaluation. These generalizations are the result of the author's study of the flying qualities, pilot workload, and pilot performance literature. They are submitted to the participants of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics to stimulate further discourse on the application of these principles to particular experimental circumstances and to encourage the readers to recognize the common threads linking the fields of aircraft flying qualities and human factors. This paper first reviews and comments on data reported by Piranian (References 1 and 2) before using these data to illustrate the general principles involved.

The Experiment. Almost ten years ago, Piranian conducted an excellent experiment on the Naval Air Development Center Dynamic Flight Simulator to determine the relative importance of some of the factors affecting the tracking performance of the pilot-airframe system. The test matrix, summarized in Figure 1, shows that the experiment was divided into two phases: buffet and flying qualities. The results of Piranian's experiment, documented in References 1 and 2, will be used to illustrate several general principles about the effects of vehicle dynamics and task difficulty on Cooper-Harper pilot opinion ratings, task performance measures and pilot workload.

Buffet Phase. The objective of the buffet phase was to quantitatively determine the effects of buffet and sustained normal acceleration on air-to-air tracking performance. The flying qualities (vehicle dynamics) were good (SAS-on) and held constant. The piloting task, as stated in Reference 1, was to track a target that took five seconds to bank into a steady-state level turn, maintained constant load factor (1.3, 2.5, or 5.0g) for 15 seconds, and returned to wings level during the final five seconds of the maneuver. Results in the buffet phase were based on tracking performance measured during the middle 15 seconds of the maneuver when the piloting task was simply to null out the tracking error left over from the initial banking segment. The target itself was not producing any motion commands during the steady state turn.

Flying Qualities Phase. The flying qualities phase of Piranian's experiment examined the effects of degraded flying qualities on tracking performance. Several sets of lateral-directional flying qualities characteristics were evaluated. The modal parameters (extracted from Reference 2) for each of these sets is shown in Table I below.

TABLE I

MODAL PARAMETERS

Parameter	SAS-on	Reference SAS-off	High-AOA	Adverse & Neutral Yaw
ζ_d	.79	.305	.408	.305
ω_d	3.59	3.24	2.13	3.24
τ_{rm}	.170	.267	.180	.267
1/T sm	.005	-.025	-.068	-.025
ζ_{sp}	.608	.608	.608	.608
ω_{sp}	5.35	5.35	5.35	5.35

In order to draw generalizations about the influence of flying qualities (vehicle dynamics) on tracking performance and pilot workload, these sets of flying qualities characteristics will be hereinafter refered to as Good (SAS-on), Marginal (SAS-off), and Poor (High-AOA). Piranian also tested "Bad" vehicle dynamics but the pilots were not able to keep the target within the visual display so tracking performance could not be quantitified.

Tracking results reported for this phase were rms tracking error and percent time within ten mils tracking error for the initial five seconds of the maneuver. During that five second period the target banked (and increased its load factor) to either 39.7 degrees (1.3g), 66.4 degrees (2.5g), or 78.5 degrees (5.0g) of bank angle. Therefore each sustained normal acceleration level reported by Piranian for this phase represents a tracking task of corresponding difficulty requiring the pilot to detect and match the direction and rate of target motion as the target performed a family of rolling and pulling maneuvers.

<u>Quantitative Results</u>. I have replotted Piranian's tracking performance results (Flying Qualities Phase) in Figures 2 and 3 with "flying qualities" and "task difficulty" as the independent variables, respectively. In Figure 2 the generally observed tendency for task performance to level-off even though flying qualities (vehicle dynamics) degrade is evident. Pilot compensation, reflected in the average pilot ratings, is responsible for this leveling-off tendency. Also suggested in this figure is the catastrophic drop-off in task performance as flying qualities degrade beyond the ability of the pilot to compensate for them. This drop-off point, where pilot capabilities are saturated, has sometimes been referred to (e.g. Reference 3) as a "flying qualities cliff".

Turning to Figure 3 we see that task performance degrades as the task becomes more difficult, which is the expected trend. It has been observed in flying qualities experiments (e.g. Reference 3) that a difficult task is generally required to expose poor flying qualities. The data in Figure 3 confirms this. There is a broad range of task difficulty where task performance differentiates good form marginal configurations but the poorest flying qualities were only exposed using the most difficult task tested. A reliable measure of pilot workload (such as pilot opinion ratings) is needed to adequately discriminate among various levels of flying qualities at the more moderate levels of task difficulty. As the task becomes quite easy no change in task performance may be detected for any level of flying qualities.

One may ask if there is the same "leveling-off" tendency in the tracking performance scores with increasing task difficulty as there was for degrading flying qualities. The results in Figure 3 are too few to tell. However they do fit the pattern shown in data taken from Reference 4 replotted in Figure 4. In that figure there is an intermediate region of task difficulty where the pilot seems to hold off some of the degradation in performance with task difficulty. At the higher levels of task difficulty, performance again falls off at a higher rate. Instead of "levelling-off" there is actually only a decrease in slope due to pilot compensation.

Cooper-Harper Scale (Reference 5). Which of Piranian's two quantitive tracking performance measures is most valid? That is, which one relates better with subjective pilot opinion? To answer these questions one must turn to the Cooper-Harper (C-H) pilot opinion ratings recorded during the experiment. The C-H pilot rating scale, shown in Figure 5, is widely used by flight test pilots such as those who participated in the Piranian experiment. It guides the pilot in making qualitative decisions about task performance and pilot workload for any compensatory piloting task. Task performance is rated as either "Desired", "Adequate", or "Not Adequate". Pilot workload, in terms of pilot compensation required during a compensatory tracking task, can be evaluated as "not a factor", "minimal", "moderate", "considerable", "extensive", or "maximum tolerable".

Figure 6 lists the pilot evaluation of task performance for each configuration (inferred from average C-H ratings) alongside the correspnding average tracking performance scores reported by Piranian. It is seen that the measurement of percent time within ten mils is more consistent with pilot opinion than rms pipper error across all levels of task difficulty. This tells us that "percent time within ten mils" corresponds much better with those aspects of performance which were of importance to those pilots in the task they understood themselves to be doing. Such a subjective validation process is recommended for any proposed quantitative performance or workload criterion using a suitable rating scale and trained pilots. The C-H rating scale may not be the best to use in this validation process because it links task performance and pilot compensation adjective pairs in a fixed hierarchy designed especially for the evaluation of airplane handling qualities. One drawback of this fixed hierarchy is that it does not permit the pilot to rate performance and workload independently (e.g. "desired" performance and "considerable" pilot compensation is not an allowable adjective pair). Any subjective performance or workload scale used to validate performance or workload criteria ought to give the pilot the freedom to do this. Such a scale could gain initial acceptance through being used alongside the C-H scale in flying qualities experiments like Piranian's. Once proven during such compensatory piloting tasks the scale would then be very useful for evaluating semi-automated tasks where pilot compensation is not the primary indicator of workload and the C-H scale is therefore of limited utility.

One additional observation should be made about the Cooper-Harper scale. It does not specifically address the different components of workload (e.g emotional stress, mental effort, multiple tasks, cognitive timesharing). Subjective workload scales that do differentiate among the various components of workload should greatly help in the validation of quantitative workload criteria and also aid the system designer. Some such scales are now under development (e.g. Subjective Workload Assessment Technique, Reference 6).

Task Performance and Pilot Workload. I have diagrammed, in Figure 7, the generally-observed trends in task performance and pilot workload as vehicle dynamics or task difficulty are varied. Since each independent variable has a similar effect on both dependent variables, vehicle dynamics (for a particular task) and task difficulty (for a particular set of vehicle dynamics) share the abscissa. Vehicle dynamics can be expressed as aircraft flying qualities (as in Reference 1), system automation (Reference 7), or display complexity (Reference 8). Task difficulty can include not only the task objectives and initial conditions (Reference 4) but also the external stressors (e.g. noise, acceleration) and disturbances (e.g. vibration, turbulence) as in References 1 and 9. The scales of Figure 7 can shift with respect to one another, depending on the circumstances and measures used; marginal vehicle-dynamics do not necessarily correspond to medium task difficulty and medium pilot workload does not presume adequate task performance. There are, of course, important factors affecting pilot workload and task performance, such as the typical variations in pilot physiology, emotional state, bias, skill and technique, which are not considered here.

Generally speaking, then, there are three regions of vehicle dynamics or task difficulty depicted in Figure 7. Region I is where the vehicle is clearly suitable for the task; performance is high and workload is low. As vehicle dynamics worsen or task difficulty increases, performance degrades steadily until Region II is entered. Here the pilot actively maintains some level of performance, or at least slows down the degradation of performance. Because of this increased pilot compensation, workload is increasing rapidly. Where the pilot's workload capacity is saturated defines Region III. Here task performance falls off a "cliff." In Region I the vehicle systems improve on what the normal pilot could do unaided; Region II is where the pilot compensates for vehicle deficiencies; in Region III the task may not be performed safely, if at all.

For vehicle or system specification purposes, Region I is more suitable for a well-defined task performance spec. However, more commonly the pacing task for system design is very challenging and usually falls, into Region II. To prevent the vehicle or system from demanding too much of the pilot, a workload specification would be preferable in that region. The workload spec could define the margin desired between the pilot workload required and the point where pilot capabilities are saturated (i.e. the Region III boundary).

Research should be done to define valid criteria to quantitatively define the axes of Figure 7. Research should then be directed at defining the Region II and Region III boundaries so that adequate specifications can be written for important mission tasks.

As performance and workload criteria are developed, experimental results should be related to the regions laid out in Figure 7. Every human factors or flying qualities experiment should consider the interrelationships between task difficulty and vehicle dynamics in the design of the experiment and during the interpretation of results. A better appreciation of these general principles should improve the value of future research and help the process of flight testing to identify pilot workload and pilot dynamics.

REFERENCES

1. Piranian, A. G., "The Effects of Sustained Acceleration, Airframe Buffet, and Aircraft Flying Qualities on Tracking Performance," Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, AFFTC-TR-82-5, Air Force Flight Test Center, Edwards AFB, California, (to be published).

2. Piranian, A. G., The Influences of Buffet, Sustained Normal Accelerations, and Basic Aircraft Flying Qualities on Tracking Performance in Air Combat Maneuvering, NADC-73257-30, Naval Air Development Center, Warminster, Pennsylvania, July 1974.

3. Smith, Rogers E., "Effects of Control System Dynamics on Fighter Approach and Landing Longitudinal Flying Qualities," Proceedings of AFFDL Flying Qualities Symposium Held at Wright State University, 12-15 September 1978, AFFDL-TR-78-171, Air Force Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio, December 1978.

4. Repperger, D. W., Rogers, D. B., Van Patten, R. E., and Frazier, J., "A Study of Task Difficulty With a Subjective Rating Scale," Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, AFFTC-TR-82-5, Air Force Flight Test Center, Edwards AFB, California, (to be published).

5. Cooper, George E. and Harper, Robert P., Jr., The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities, NASA TN D-5153, April 1969.

6. Reid, Gary B., Eggemeier, Thomas, and Shingledecker, Clark A., "Subjective Workload Assessment Technique," Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, AFFTC-TR-82-5, Air Force Flight Test Center, Edwards AFB, California, (to be published).

7. Cord, Thomas J., IFFC Flying Qualities Simulation, AIAA-81-1821-CP, AIAA Guidance and Control Conference, Albuequerque, New Mexico, August 19-21, 1981.

8. Korn, Jonathan, Gully, Sol W., and Kleinman, David L., "Modelling, Design and Validation of an Advanced Cockpit Display System," Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, AFFTC-TR-82-5, Air Force Flight Test Center, Edwards AFB, California, (to be published).

9. Nardi, Louis U., Kawana, Harry Y., and Greek, David C., Ride Qualities Criteria Validation/Pilot Performance Study-Flight Test Results, NASA Contractor Report 144885, September 1979.

TEST MATRIX

BUFFET PHASE

FLYING QUALITIES: STABILITY AUGMENTATION SYSTEM ON

BUFFET: 0 TO ±.5 GZ AND GY (10 CPS)

ACCELERATION: 1.3 TO 5 GZ

6 PILOTS X 4 SETS X 12 RUNS = 288 RUNS
(24 RUNS PER DATA POINT)

FLYING QUALITIES PHASE

FLYING QUALITIES: STABILITY AUGMENTATION SYSTEM OFF
HIGH - AOA STABILITY DERIVATIVES

AILERON YAW: PROVERSE
ZERO
ADVERSE

BUFFET: 0 AND ±.25 GZ AND GY (10 CPS)

ACCELERATION: 1.3, 2.5, 5.0 GZ

5 PILOTS X (3 SETS X 2 F.Q. + 2 SETS X 3 A.Y.) = 60 RUNS

Figure 1. Test Matrix of Piranian Experiment

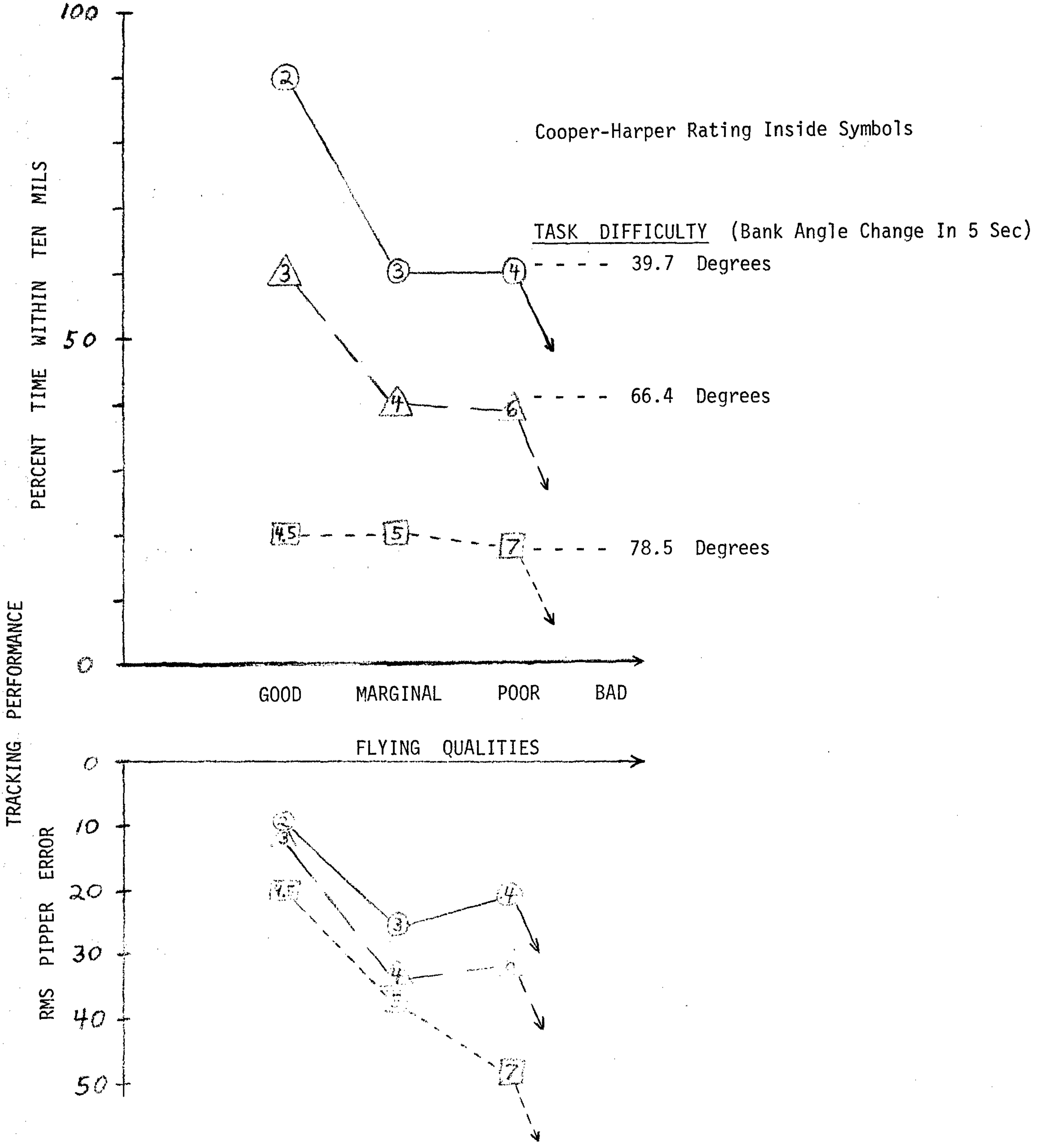


Figure 2. Tracking Performance Versus Flying Qualities (Piranian Data)

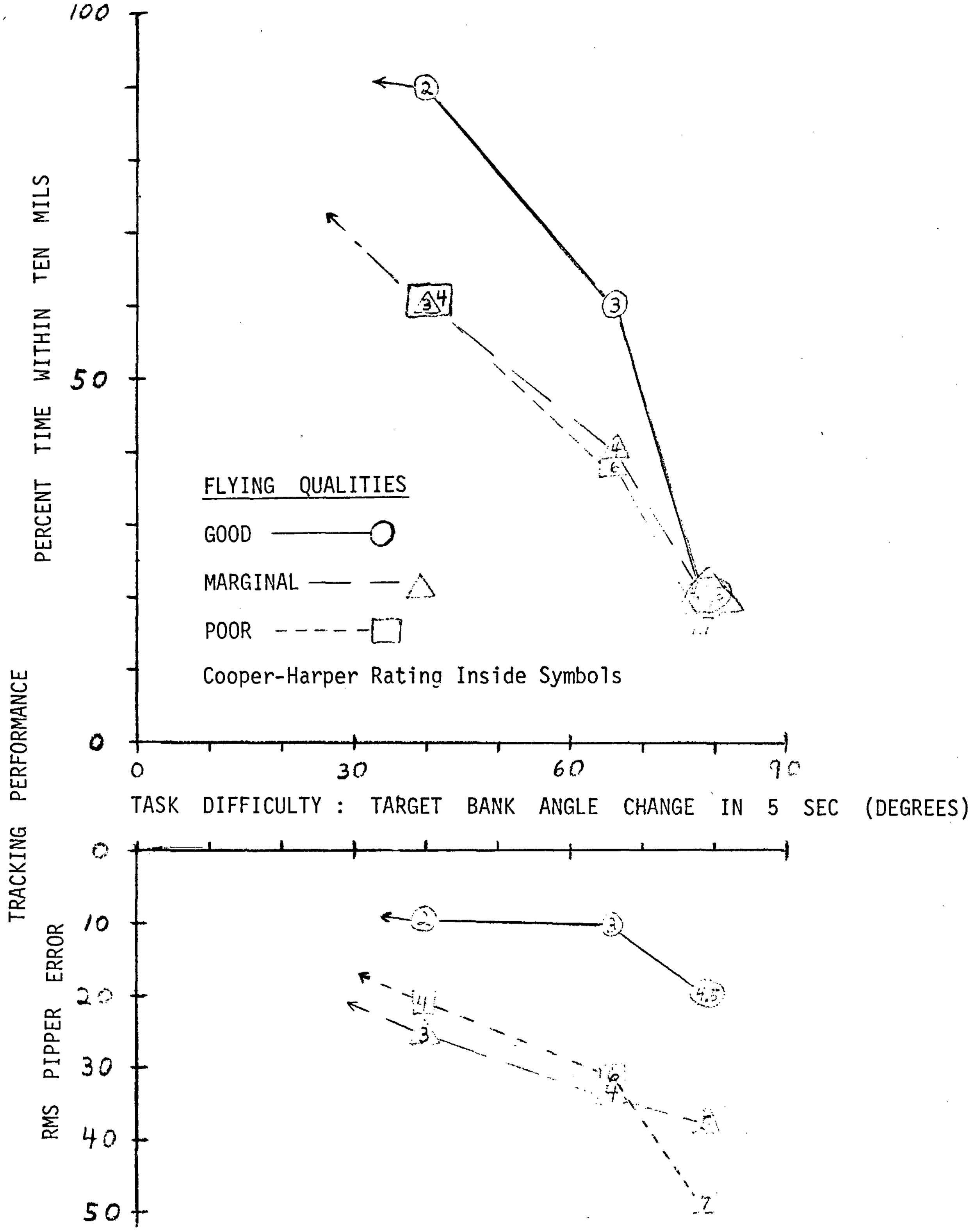


Figure 3. Tracking Performance Versus Task Difficulty (Piranian Data)

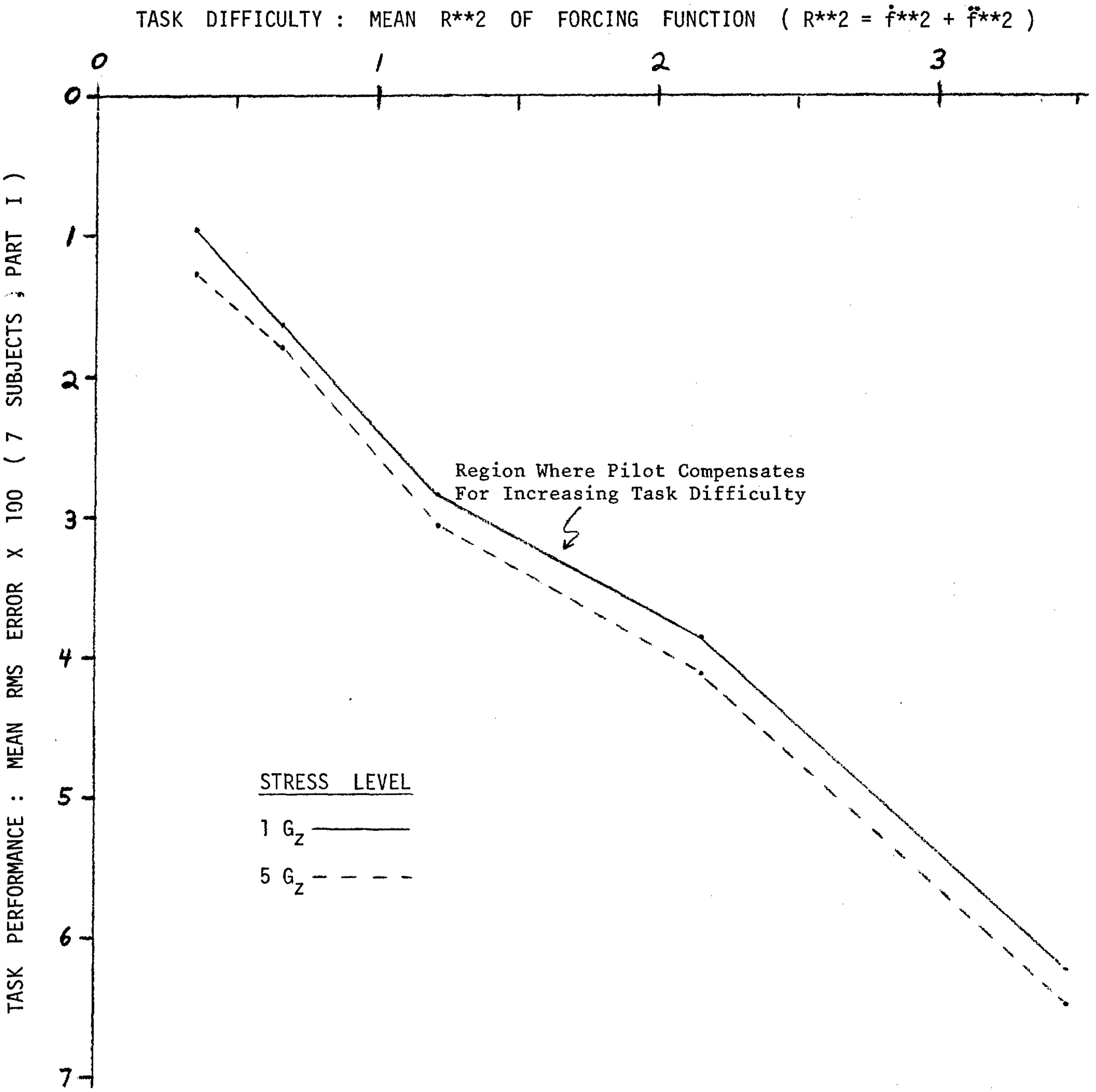
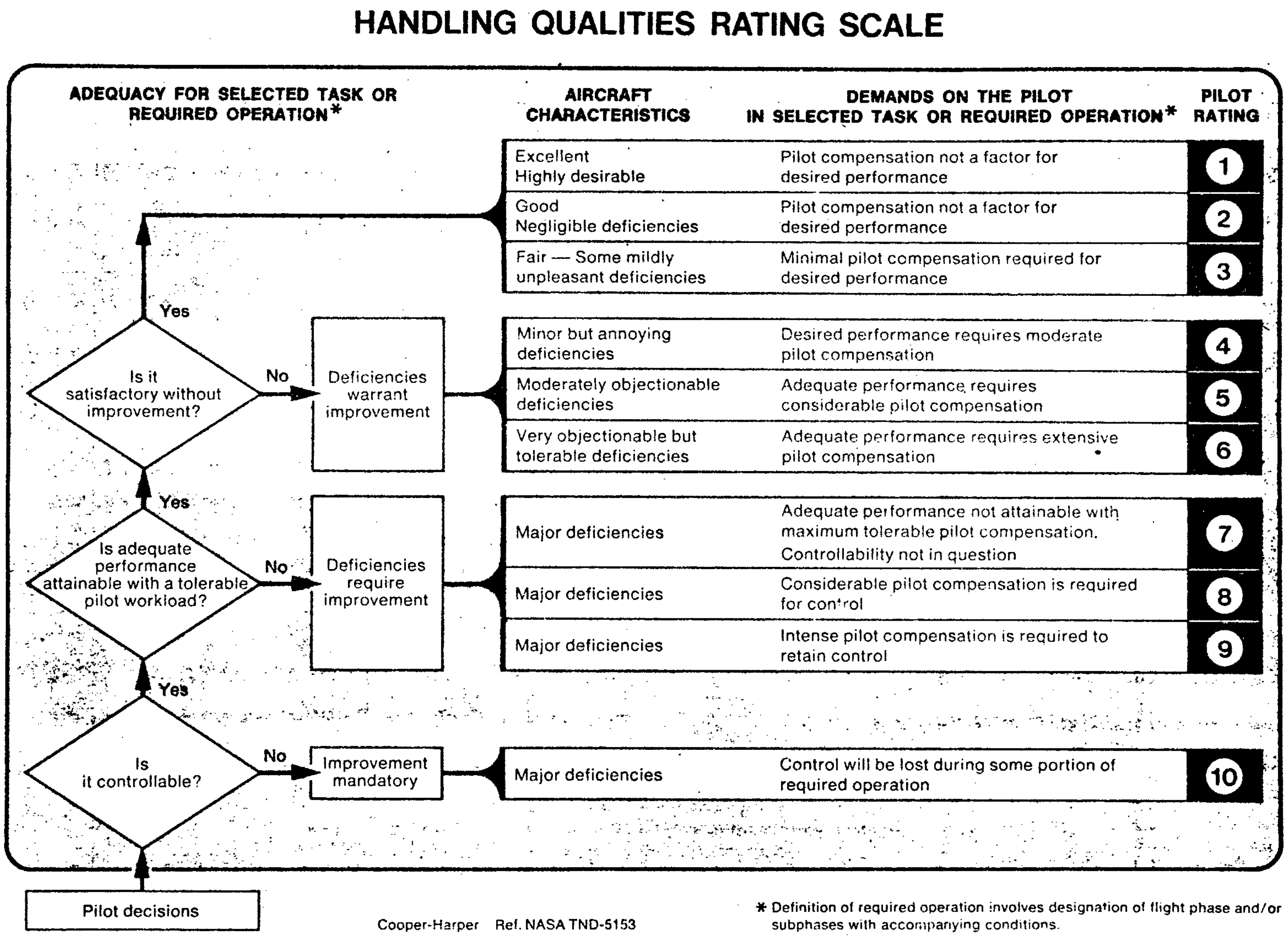


Figure 4. Tracking Performance Versus Task Difficulty (Repperger Data)



Cooper-Harper Ref. NASA TND-5153

* Definition of required operation involves designation of flight phase and/or subphases with accompanying conditions.

Figure 5. Cooper-Harper Pilot Rating Scale

VALIDATION OF QUANTITATIVE PERFORMANCE MEASURES

RMS PIPPER ERROR			
9	D		
11	D		
20		A	
21	D		
26	D		
32		A	
33	D		
37		A	
48			N

% TIME WITHIN 10 MILS			
18			N
20		A	
20		A	
39		A	
40	D		
60	D		
60	D		
60	D		
90	D		

D= DESIRED PERFORMANCE (C-H 1-4)

A= ADEQUATE PERFORMANCE (C-H 5-6)

N= NOT ADEQUATE PERFORMANCE (C-H 7+)

Figure 6. Tracking Performance With Pilot Opinion

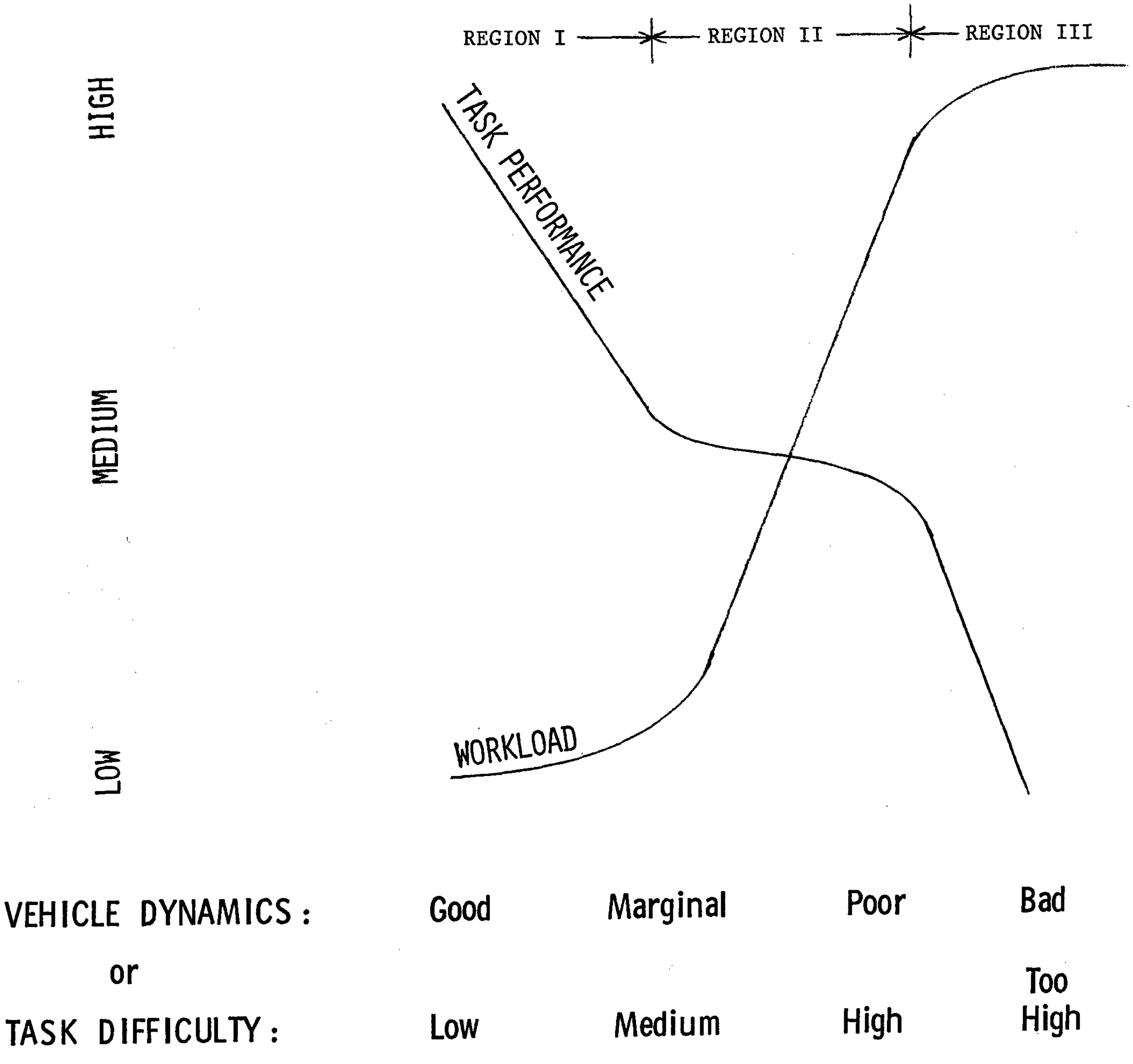


Figure 7. Task Performance And Pilot Workload Versus Vehicle Dynamics Or Task Difficulty

AUTOMATED AIRCRAFT AND FLIGHT CONTROL SYSTEM DESIGN

by

Morris A. Ostgaard
Air Force Flight Dynamics Laboratory
Air Force Systems Command
United States Air Force

Automated aircraft and flight control design has been a subject of discussion and research and development since the early 1900's. The key issue at this early date was pilot workload and the application of automation or feedback control to solve workload was a subject of research and experimental flying. The first application of guidance and control to aircraft was Dr Elmer Sperry's automatic stabilizer, later called auto pilot, for which he received a 50,000 French franc prize in 1914 for the most stable and safest aircraft. His main purpose in developing this system was to provide a mechanism to relieve the pilot from the detailed task of continually stabilizing the vehicle and thereby making it a useful machine. Figure 1 is a functional diagram of this system which illustrates some of the basic feedback principles involved. To provide some background for later discussions, the following will discuss some of the characteristics of the human operator and their similarity to guidance and control functions (Reference 1). The operator's characteristics as a controller depend on four kinds of variables: control task variables, which include the system inputs and all the system elements external to the operator; environmental variables such as ambient illumination, temperature, vibration, etc.; operator-centered variables such as training, fatigue motivation; and procedural variables such as instructions, practice, order of presentation relating to a given task. When these variables are essentially time stationary or invariant over an interval of interest, the operator vehicle system can be modeled as a quasi-linear system much the same as standard servo loops. Figure 2 tends to illustrate some of these relationships showing that the major part of the human operator problem is really one of observability and controllability, and the fact that one of his major tasks is the integration of all the information available and attempting to determine the strategy or control aspects that best satisfies the objective of the mission task to be achieved.

The fundamental aspects of manual vehicular control theory can be briefly summarized as follows:

a. To accomplish guidance and control functions, such as flying a desired track in the presence of disturbances, maintaining position control as in formation, terrain following or refueling, and flying intercepts or approaches, the human operator structures a variety of closed loops about the vehicle to achieve the control actions and functions desired in terms of actual vehicle motions and flight path.

b. To be satisfactory, these closed loop systems, which include both active and passive elements, must share certain of the dynamic characteristics of a good closed-loop control and stablization system. As the adaptive element to accomplish this end, the pilot must make up for any dynamic deficiency of the system by appropriate adjustment of its dynamic properties. This is the terminology that was referenced as integration in the form of Figure 2.

c. The cost of this adjustment is fundamentally workload and can consist of stress, which includes vibration, hazards, and environment, concentration of operator faculties on the mission task dynamics, and the capability to cope with the unexpected. This, then, illustrates why there is a close relationship between guidance and control and pilot workload in that the guidance and control systems can be implemented and designed to reduce many of the normal pilot operator functions as well as defining the true information or states required for the pilot to execute his mission.

To illustrate some of the trends and history of guidance and control over the past few decades, Figure 3 illustrates briefly what these guidance and control technology trends are. It summarizes some of the transitions that have occurred over the past two and a half decades, beginning with the 1950s (Reference 2). But most significantly, it illustrates two basic trends: the first is the increase in guidance and control functions and requirements; the second is the progression from vehicle to weapon systems into what we now call air warfare systems. During the early 1950s, the trend was predominately vehicle oriented in terms of stability augmentation systems to improve the response in control characteristics. This was followed by closed-loop terrain following systems to improve adverse weather, low altitude performance in safety aspects of the system. Adaptive system concepts became operational in the F-111. These concepts, which have been discussed extensively in other publications, contributed substantially to the current modern control theory and technology. As a sideline, a seldom recognized fact is that the Kalman filtering theory was a direct outgrowth of the theoretical work undertaken to describe, analyze, and design adaptive systems.

Emergence of fly-by-wire technology in various forms became evident during the early 1960s with fly-by-wire spoiler controls on the F-111 and the electronic flight control on the French MIRAGE. Increased use of inertial sensors and navigation systems to augment vehicle of flight path control for reducing landing minimums, and improve interaction with area navigation and advanced air traffic control concepts during the early 1960s, represents initial steps in the functional integration for increased performance and workload reduction. During the mid-1960s, the benefits of feedback control could offer in alleviating structural fatigue problems and turbulence sensitivity were recognized. This work culminated in the application of maneuver load control and structure load control to alleviate turbulence induced fatigue problems, and improve crew ride quality in the B-1 and other systems. Progressing into the early 1970s, the MRCA, the Swedish Viegin, French Concorde, the F-16, and YC-14 designs witnessed incorporation of sidestick controllers, electronic displays, increased emphasis on digital techniques, control and propulsion system dynamic interaction, direct lift control, ride control, and relaxed static stability, all of which were made possible through implementation of emerging technology.

In looking toward the future of air warfare systems, the impact that command and control and communications will have on the vehicle guidance and control in terms of tactical control is significant. This need generates another guidance and control dynamics loop which dictates substantial dynamic and functional interaction to provide the desired operational capability. The time-space positioning capability permitted through accurate position fixing, employing advanced navigation systems, and the onboard vehicle trajectory control permits the implementation of the command and control function to marshal forces for tactical deployment and, also, offers a means for redirect capabilities.

The six-degree-of-freedom control projected through active control technology permits design freedom and tactical capabilities unavailable heretofore other than in specialized rotary wing configurations. These capabilities have stimulated application of modern control theories involving strategies, differential gaming, to determine means that can provide optimum trajectories and tactical options to increase weapon delivery accuracy and minimize pilot workload.

With this brief overview in projection, let's examine advances in guidance and control to determine how they can be employed to reduce pilot workload, particularly in the low altitude, high-speed regime. A considerable amount of work has been performed on the application of stabilization systems to reduce pilot workload. However, the majority of these systems have not reduced the pilot workload to any large degree because, in most cases, they introduced unconventional methods of control which, in themselves, created a learning problem (Reference 3). The learning of these new control techniques aggrevated, in many cases, the workload problem. Also, many of these developments concentrated primarily on the control aspects ignoring the information display aspects which provides the necessary information for the pilot to act as a closed-loop controller.

The problem associated with combining guidance and control systems or automatic systems with pilot operation is the mechanism of pilot interaction. This is somewhat analogous to the cruise control on an automobile which works perfectly for cruising on a highway; its use in traffic, however, involves a tremendous amount of switching and other unconventional manipulation. In terms of control, the first attempt to solve some of these problems was control wheel steering wherein a switching or sensor mechanism on the wheel or stick was employed to sense applied force by the pilot in a normal manner, deactivating some of the stabilization modes of the system and, thus, restore normal aircraft maneuvering response. This proved successful to a degree but the fundamental control laws involved were for stabilization, and the limited authority of the system made it extremely restrictive and added to the pilot workload rather than easing it. A more successful approach was the command augmentation system design. This is illustrated in Figure 4 and shows the requirements and characteristics involved, the typical representation of the system and, finally, the basic issues involved. The command augmentation system provided a means for enhancing the stabilization properties of the vehicle without deteriorating the maneuvering performance while at the same time provided an input and control law capability that made the vehicle behave in a manner that was very desirable from the handling qualities standpoint. This was successful for the traditional up-and-away and some precision tasks such as approach and landing; however, for certain weapon delivery or mission functions, the general control laws were proven inadequate. Further control analysis and flight research experience indicated that an approach termed task-oriented control laws wherein the control characteristics of the vehicle are deliberately configured to provide precision control and performance for that particular mission would be a solution. These have eliminated many of the problems of integrating the pilot with the control system and are proving very successful, and indicate that the vehicle control aspect of the problem is solvable employing guidance and control design technology. At this point it may be advantageous to examine the role of tactical air focus and how this may affect the design of the flight control system, its role in system integration and related areas of automation.

The primary role of the tactical Air Forces, because of their mobility and concentrated fire power capability is stabilization of theatre conflict as illustrated in Figure 5. In the ideal case prior to ground force engagement, the tactical Air Forces can reduce the Pact ground force level to a point where ground forces alone can successfully defend and contain. In a minimum case for multiple engagements, the tactical air and ground forces can reduce the Pact of ground force level to a point where tactical forces together can successfully defend. In any event since a cooperative effort between the surface and Air Forces is necesarry, secure communication, external threat data and time-space positioning become vital elements for tactical battlefield control. Figure 6 through 11 represent a scenario of events as the engagement progresses. The key situation to avoid, as illustrated in Figure 11, is a breakthrough which demands a higher level of both air and surface forces to counter and stop the primary exploitation force. Since enemy defenses and weather can have a significant effect on the ability of the tactical Air Forces, Figure 12 illustrates projected Soviet air defense threats at the FEBA. These threats are considered to be extremely mobile and can follow the initial forces to the FEBA. If adverse weather conditions are super-imposed upon this threat, the only tactical option available is a low altitude penetration and attack to counter the advancing forces.

Since this role of the tactical forces when combined with the adverse weather and high threat environment creates a high crew workload, a study was performed on the impact of automation on crew effectiveness and efficiency. Figure 13 illustrates the approach taken to assure a numerical methodoloy. This methodology is divided into two sections, a requirements section and a functional/means section. One basic requirement is stabilization of theatre conflict, which is the principle role of the tactical Air Force. Consistent with this role is the necessity to improve effectiveness and avoid catastrophy. Catastrophy is defined as a loss of aircraft and crew members due to battle damage or on-board critical failure.

A means or functional capability to achieve these ends or requirements are illustrated in the second section, but not necessarily in the order of importance for the various mission segments. This approach provides a mechanism whereby functional interactions during various mission segments can be assigned numerical values for determining areas of importance.

Figure 14 illustrates a functional approach summary where the previous functions and means are organized in terms of a basic mission control function, mission operational functions and mission support functions. The fundamental mission control function is flight path or trajectory control, which is the only mechanism available for controlling the vehicles direction and magnitude of the velocity vector to achieve survivable penetration and timely terminal conditions for the weapon delivery. The principle role of the crew is that of managing and operating the mission operational functions to achieve the mission control function. These mission operational functions vary according to the mission segment, as shown by the ingress column, which includes flight control, propulsion control, navigation, target acquisition and selection, threat warning, fire control, and IFFN. To provide this capability a number of mission support functions are required. These are electrical distribution systems, hydraulic distribution systems, a data distribution network, warning systems, and life support which includes crew escape. This structure or functional approach, indicates that the principle role of the pilot or crew is that of functional and information integration as opposed to vehicle control.

With this approach as a functional baseline, the sources of pilot workload, which is one of the critical aspects of the crew station was addressed. Figure 15 lists the five major sources of workload, these are: input saturation, multi-tasking, time line compression, pilot bandwidth limitations, and fatigue aspects; such as continuous low level operation. Using these sources of workload, an approach to define guidelines for which of these functions should be automated and how should the functions be integrated was formulated. Figure 16 illustrates one approach used in examining an air-to-target sensing and acquisition problem. The task performance requirements, such as time, error, training, and crew loading are identified and the various functional crew aspects such as sensory, motor, cognitive, and fatigue. The descriptors for coding and cross hatching for each of these task performance requirements are shown at the bottom of the figure, with some guideline criteria for automation; such as, automate routine functions, automate memorization tasks, automate precision operations, automate sequentially timed operations, but do not automate judgmental aspects nor tactics as the structure of the battle varies. To provide some insight into this particular area of the mission, Figure 17 indicates some of the crew interactions that can occur during ingress, engagement and egress phases of an air-to-ground mission. These are the same means and functions listed in Figure 10. The numerical values at the bottom of each metric indicates the number of visual and crew motor interactions. The top three interactions indentified are: threat warning and countermeasures versus flight control, propulsion control and flight control, and navigation and flight control. with the top six systems being flight control, threat warning, navigation, target sensing, external data and weapon delivery.

With this insight into the number of interactions and the level of multitasking that can occur as a function of the mission segments, Figure 18 summarizes the time line analysis for this same air-to-ground mission. The mission is broken down into ten second segments for ingress engage, and egress to illustrate the level of simultaneous tasking and the relative importance of each of these functions.

Since this multi-tasking problem is emerging as the principle source of crew workload, it was necessary to consider what are potential areas of automation and what has prior experience shown in this area. Figure 19 is a summary of candidate criteria in areas of automation based upon a paper presented by MIT in the early 60's, a recent literature survey and test experience.

The commonality of these candidate criteria is not suprising and support earlier guidelines to automate highly repetitive functions, which are essentially regulatory functions in characteristics and difficult functions, which can be either high order or high bandwidth functions and time critical failure functions. TAC experience indicates similar areas that could be candidates for automation.

The next series of Figures, 20 through 29, are technology oriented and provide a description of the functions listed in Figure 14, and detailed discussion of the current approach to automation, what is an automation approach, what are technologies needed for automation, what are some current programs that address the needed technology, and finally, what are some new thrusts or emphasis required to better achieve this capability. Since many of these charts are self evident, very little discussion is required other than to note that flight path control is dominant throughout all the mission phases with threat warning and countermeasures as supporting technologies during the ingress, engagement and recovery phases. This technology breakdown and discussion of the interacting aspects of many of these functions suggests there are distinct areas of commonality between automation and integration and a role for each.

The final phase of the study activity was devoted to identifying a total system integration approach and relate automation to integration.

Figure 30 shows the structural integration/automation approach which is essential for automation in combat aircraft. Figure 31 indicates that high levels of automation is currently incorporated in combat aircraft but limited to the specific subsystem areas; for example, fire control is highly automated, flight control is highly automated, many of the weapon controls and threat warning functions are highly automated including sensors. The most probable reason for this is that specific functional bottlenecks are attacked as problems occur, the logical extension of existing designs, the influence of organizational alignments and limited cross technology integration technology.

Because of the cross technology integration limitation, Figure 32 illustrates that the pilot is now the automation core among the various functional elements. He performs the functional integration of information from multiple sensors with the flight control, navigation, threat warning, fire control, propulsion control, weapons and displays, and executes the control functions. Since the multi-tasking problem is one of the largest areas that can influence crew workload and place a practical outer limit on performance capabilities, a potential solution to this problem is shown in Figure 33. The solution is to re-allocate functions by implementing a different automation core. A logical core is the flight trajectory and attitude control functions, because its a common function, is dominant in a mission control, is central to any air vehicle activity and encompasses several key functional areas, fire control, navigation, flight control and propulsion.

Figure 34 illustrates a flow diagram from the logical automation progression from present day levels of automation; where the crew integrates all the functions, to the core automation approach; where navigation, flight control propulsion, and fire control are the core functions; where the crew only integrates the threat weapons and sensors information; and finally, the ultimate level of automation; where the crew operates as an overall management function since automation and integration is implicit in the overall design. Figure 35 depicts a conceptual core approach and development which is essential to continued automation. The input to the core function are sensors, threats, external data with the output being control, propulsion, weapons and display. Because it is a universal interface core, it's a basis for input/output specification for common equipments and common software elements and it's a logical growth to higher automation levels as input hardware improves. The alternative to this approach is proliferation of automation approach and hardware. This automation core is a flexible system structure involving logical partitioning and interface standards as shown in Figure 36, that can be used as a basis for developing a prescribed system or implementation architecture consisting of distributed or concentrated redundant processors, determination of firmware or hardware, determination of various bus structures; and finally, standards applicable to multifunctional areas.

The future trends tend to indicate the need for increased emphasis on precision flight path control, increased emphasis in flight control, propulsion, navigation, and fire control interaction and integration as a core structure; and finally, the increase need for a common data distribution network as shown in Figure 38, which stresses integrity and capacity. Figure 39 illustrates a potential functional flow diagram which identifies a functional core to the right of the diagram, which encompasses flight control, navigation, propulsion, and weapon delivery as key elements to flight path control. With the inputs being fire control based upon target sensing and acquisition information, threat warning based upon information from the on-board and external sources; and finally, external data for communication, threat updates and marshalling. The major challenge to achieve flight path or trajectory control for the future is the development of a core architecture and structure with the capacity, integrity, and flexibility to support these trends. Figure 40 depicts a conceptual diagram of possible system design for the future, indicating that control theory and computation is dominant in defining the system mathematical and functional structure, which then can effectively be implemented in a data distribution and information processing system.

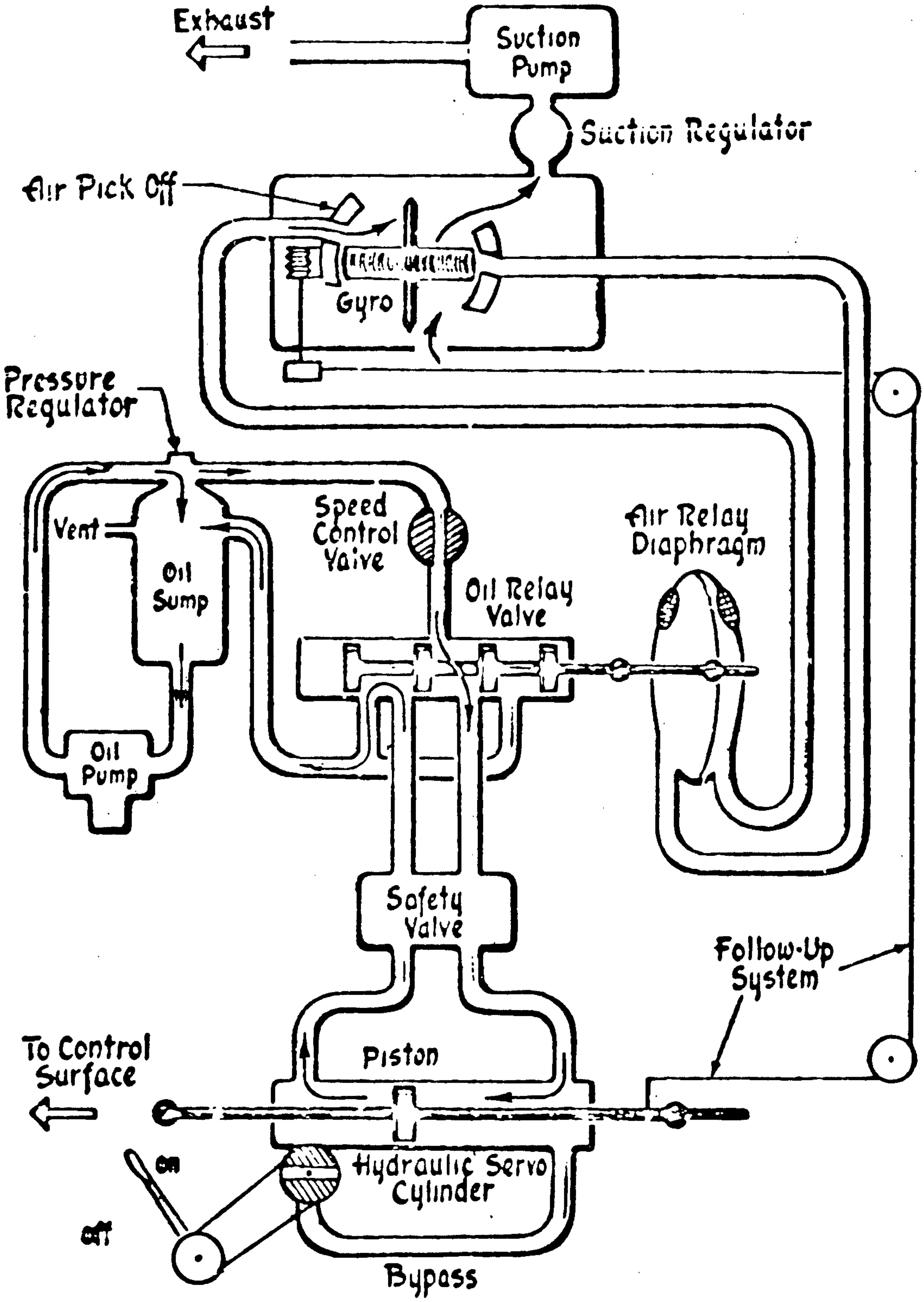


Figure I-1 Early Sperry Autopilot

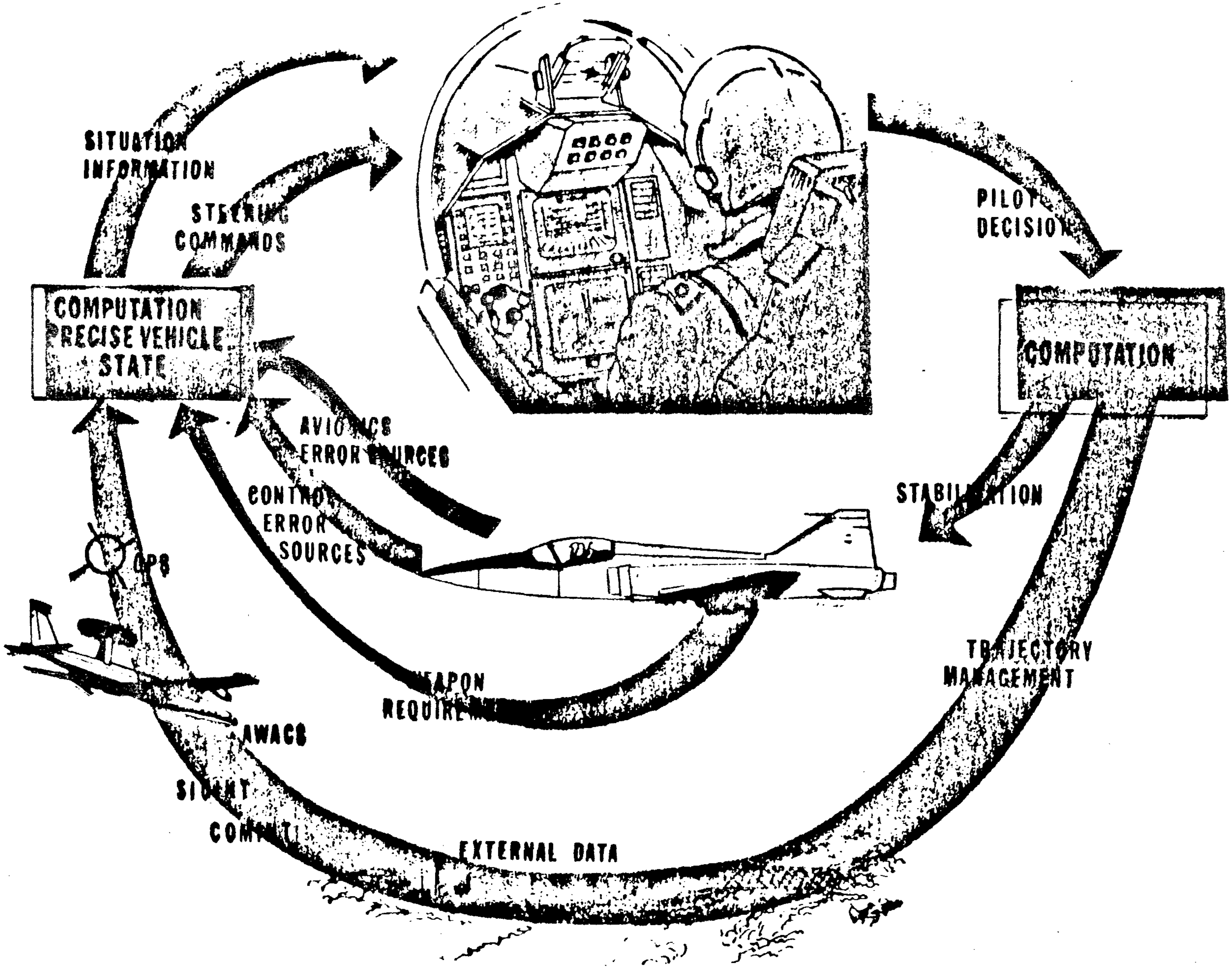


FIGURE 2: SIGHT INFORMATION

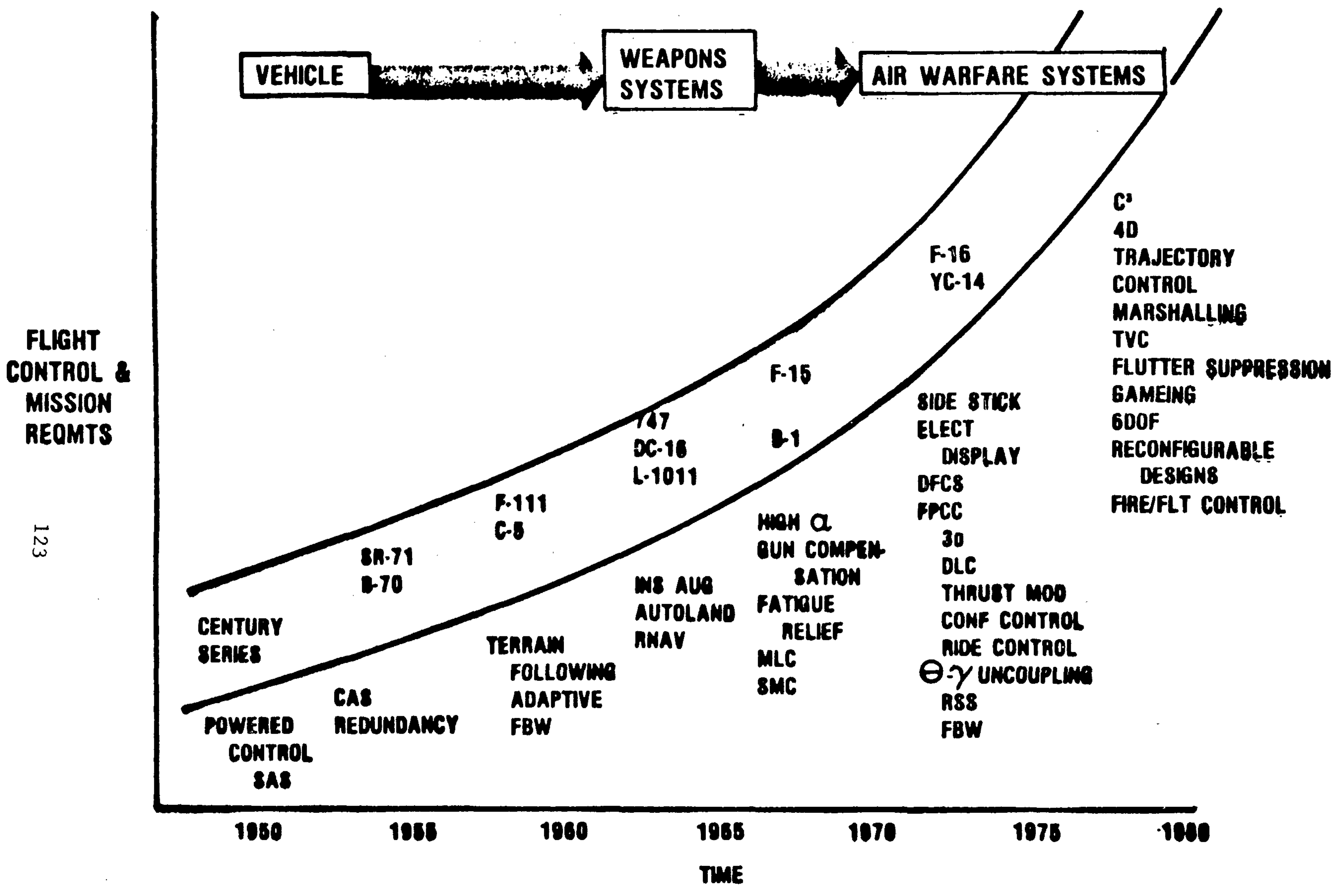


FIGURE 3: CONTROL TECHNOLOGY TRENDS

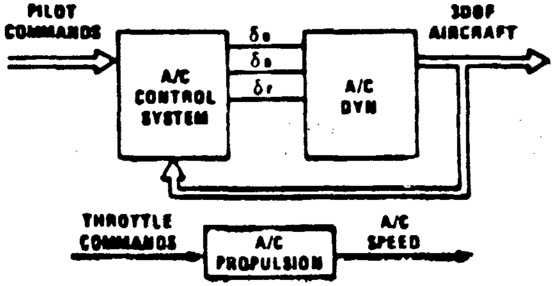
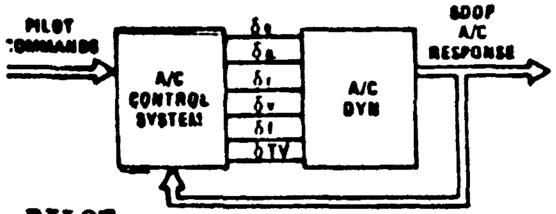
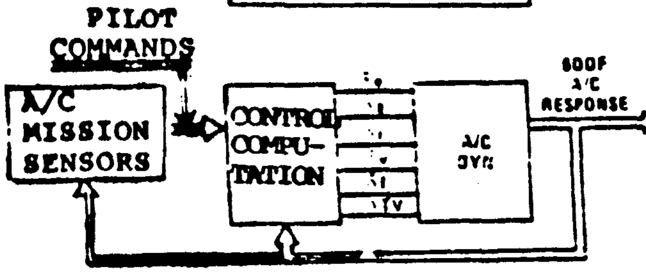
APPLICATION	CURRENT REQUIREMENTS	CHARACTERIZATION	DESIGN & ANALYSIS	RELIABILITY ISSUES	COMPUTER IMPLEMENTATIONS
CCV	STABLE AND CONTROLLABLE CLOSED LOOP AIRCRAFT SYSTEM		CLASSICAL STABILITY ANALYSIS TECHNIQUES, MULTIVARIABLE CONTROL TECHNIQUES??, TRANSIENT RESPONSE ANALYSIS-SIMULATION	[FAIL OP]^N FAIL SAFE ELECTRONICS SOFTWARE	M REDUNDANT CENTRALIZED COMPUTERS
CCV + 6DOF	+ IMPROVED MANEUVERABILITY CAPABILITY		MULTIVARIABLE CONTROL TECHNIQUES?	MORE COMPLEX REQUIREMENTS	MORE COMPLEX SOFTWARE
INTEGRATED CONTROL	INTEGRATE WITH GUIDANCE FIRE CONTROL PROPULSION CONTROL WEAPON DELIVERY STRUCTURAL CONTROL FOR IMPROVED PERFORMANCE AND RELIABILITY AT LOWER POWER, SIZE, WEIGHT AND COST		MULTIVARIABLE CONTROL TECHNIQUES MULTIVARIABLE OPTIMIZATION TECHNIQUES	RECONFIGURABLE FAULT-TOLERANT SOFTWARE/HARDWARE	DISTRIBUTED MICROPROCESSOR ARCHITECTURE

FIGURE 4: AIRCRAFT CONTROL SYSTEMS - CURRENT EXPERIMENTAL SYSTEMS AND PROJECTED

PRIOR TO GROUND FORCE ENGAGEMENT -

TACTICAL AIR FORCES REDUCE PACE
GROUND FORCE LEVEL TO POINT WHERE
SURFACE FORCES ALONE CAN SUCCESSFULLY
DEFEND

IDEAL CASE

PRIOR TO GROUND FORCE ENGAGEMENT -

TACTICAL AIR FORCES REDUCE PACT
GROUND FORCE LEVEL TO POINT WHERE
SURFACE FORCES AND TACTICAL AIR FORCES
TOGETHER CAN SUCCESSFULLY DEFEND

MINIMUM CASE

SURFACE AND AIR FORCES TOGETHER MUST --

- SEE THE BATTLEFIELD - VISUAL OR SYNTHETIC
- CONCENTRATE COMBAT POWER AT CRITICAL TIMES AND PLACES
- FIGHT AS A TEAM

AIR/LAND BATTLE REQUIREMENTS

FIGURE 5

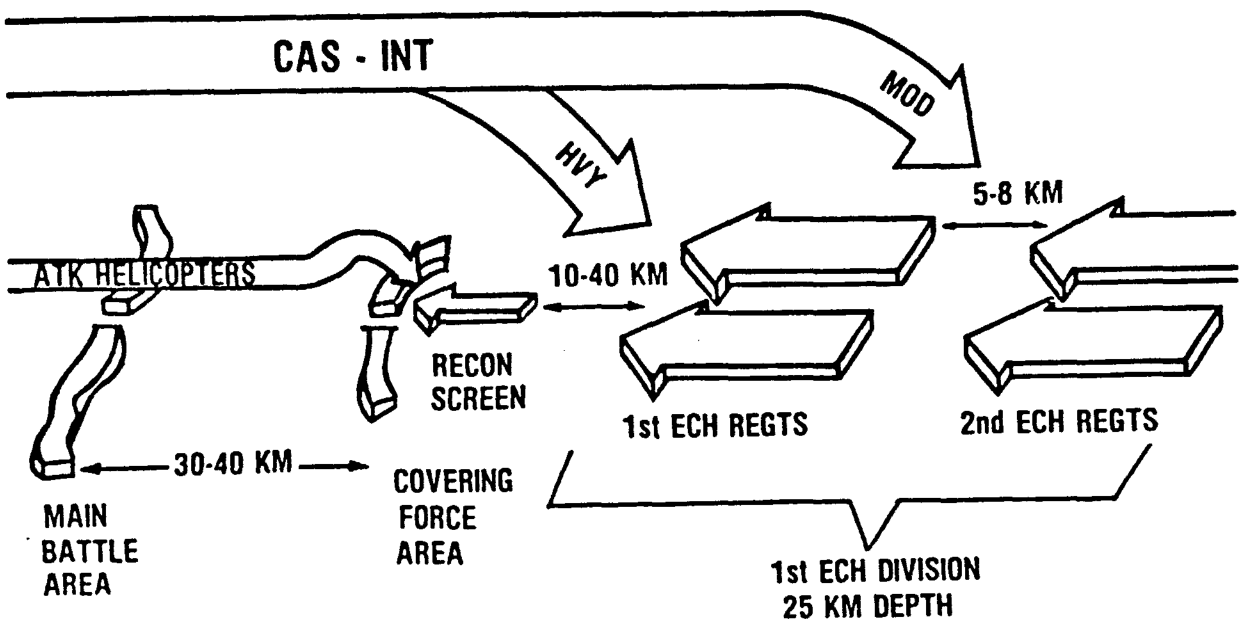


FIGURE 6 INITIAL CONTACT

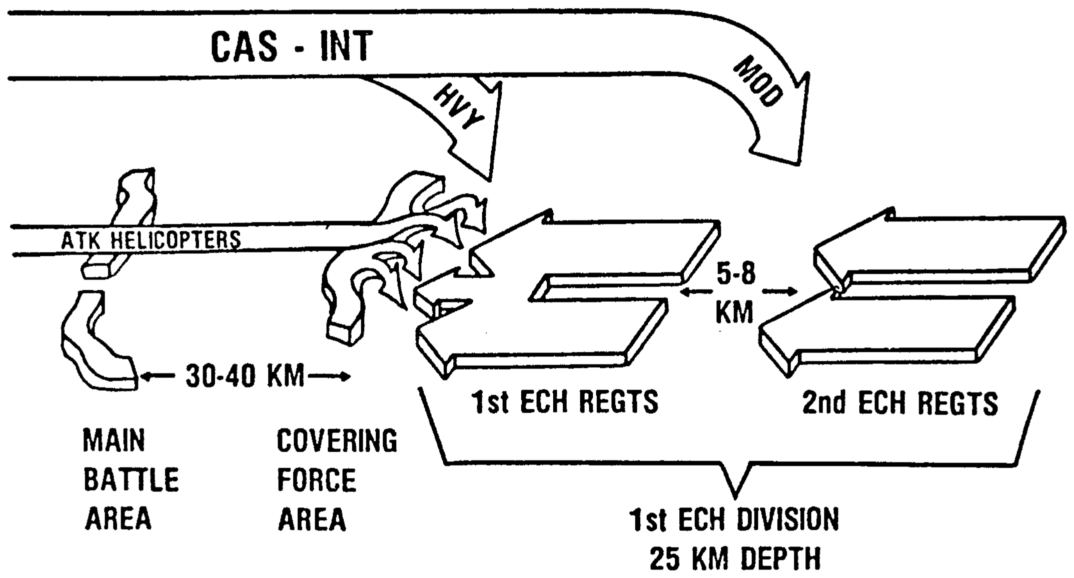


FIGURE 7 CLOSURE OF LEAD REGIMENTS

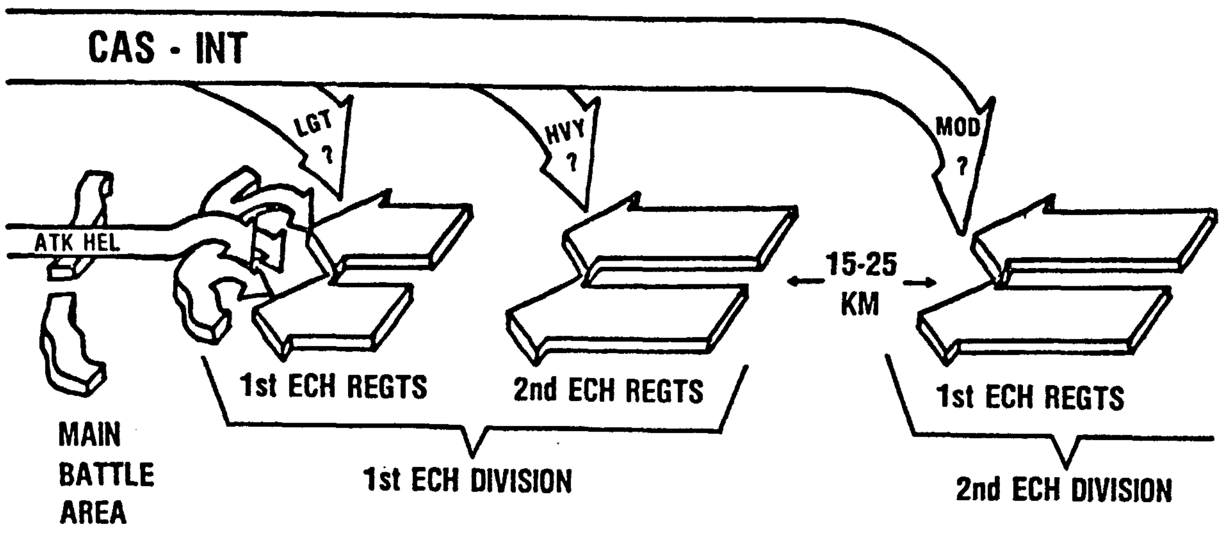


FIGURE 8 COVERING FORCE DELAY

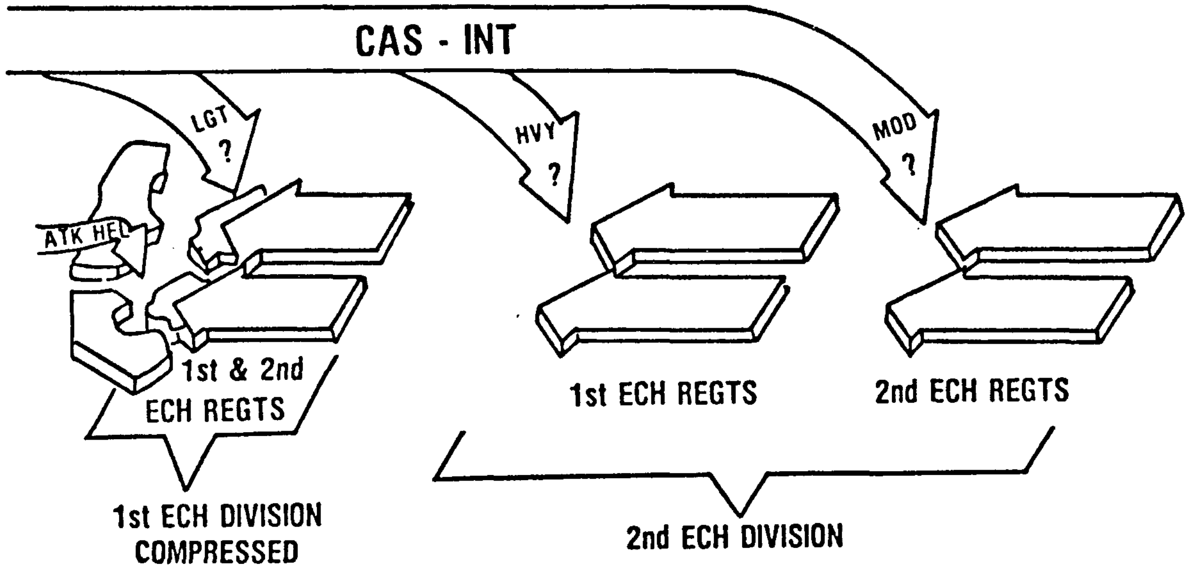


FIGURE 9 ENGAGEMENT - MAIN BATTLE AREA

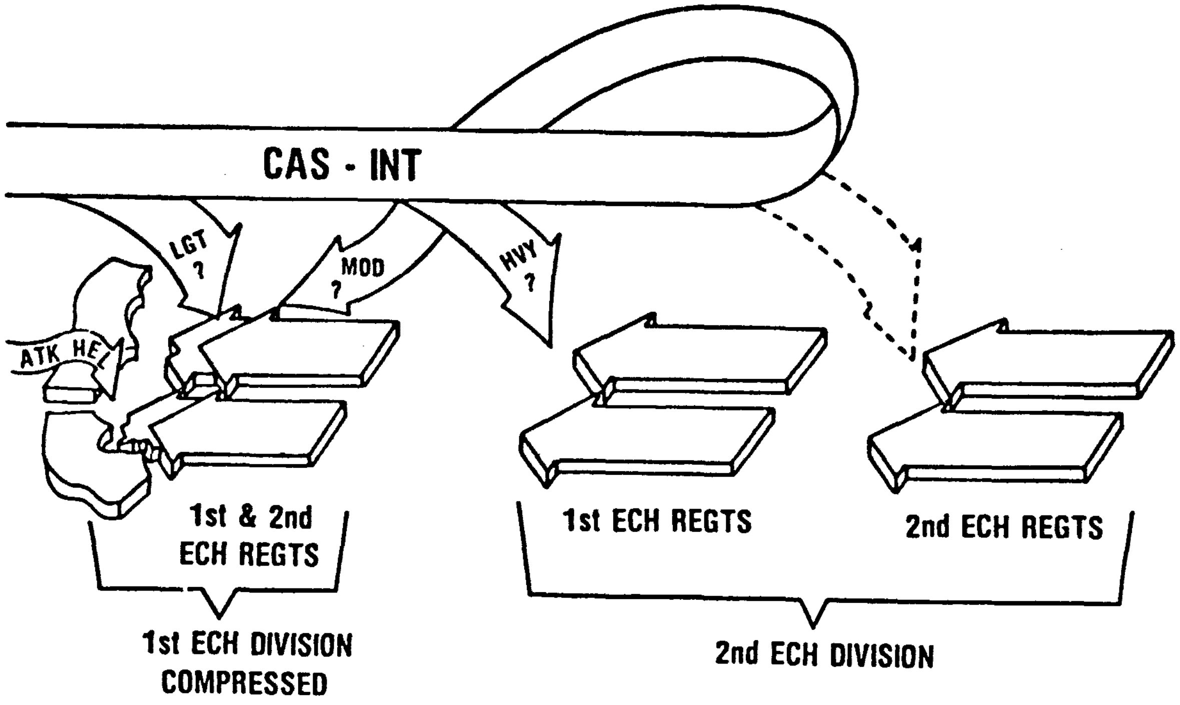


FIGURE 10 REDISTRIBUTION OF EFFORT

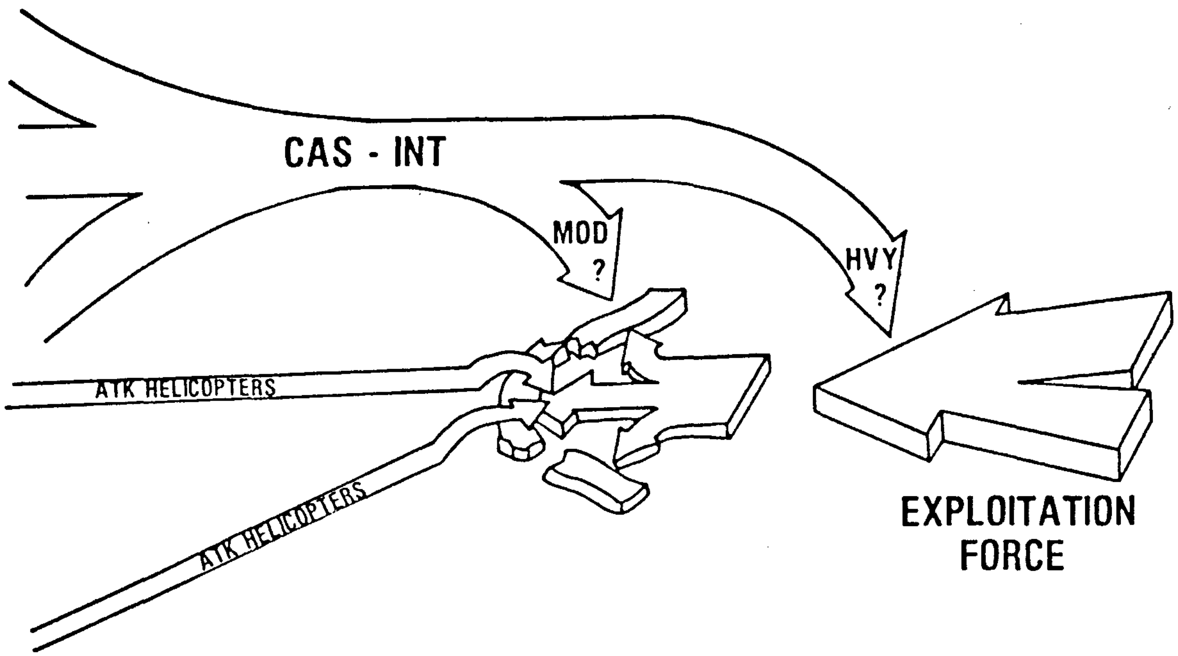


FIGURE 11 BREAKTHROUGH

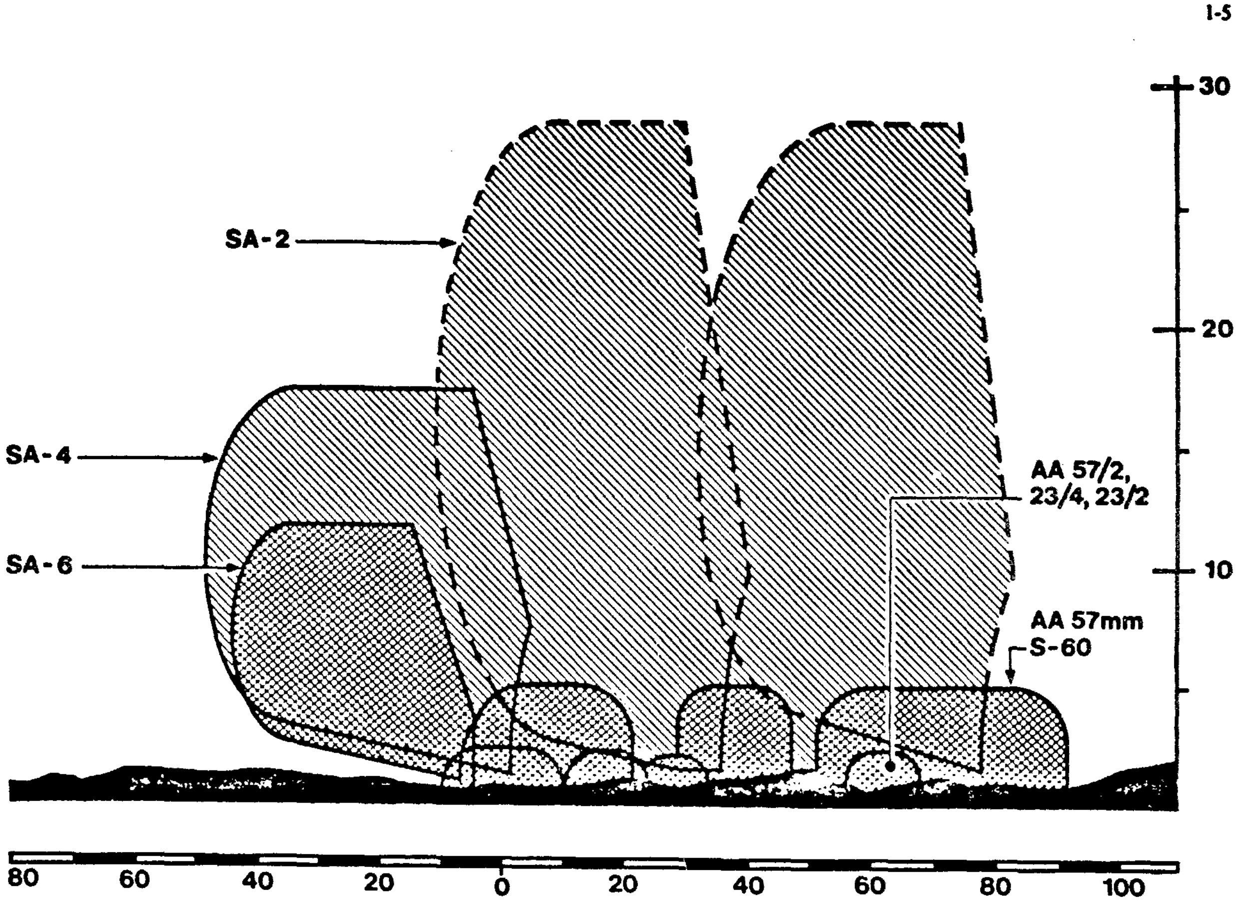


FIGURE 12 POSSIBLE SOVIET AIR DEFENCE THREAT AT THE FEBA

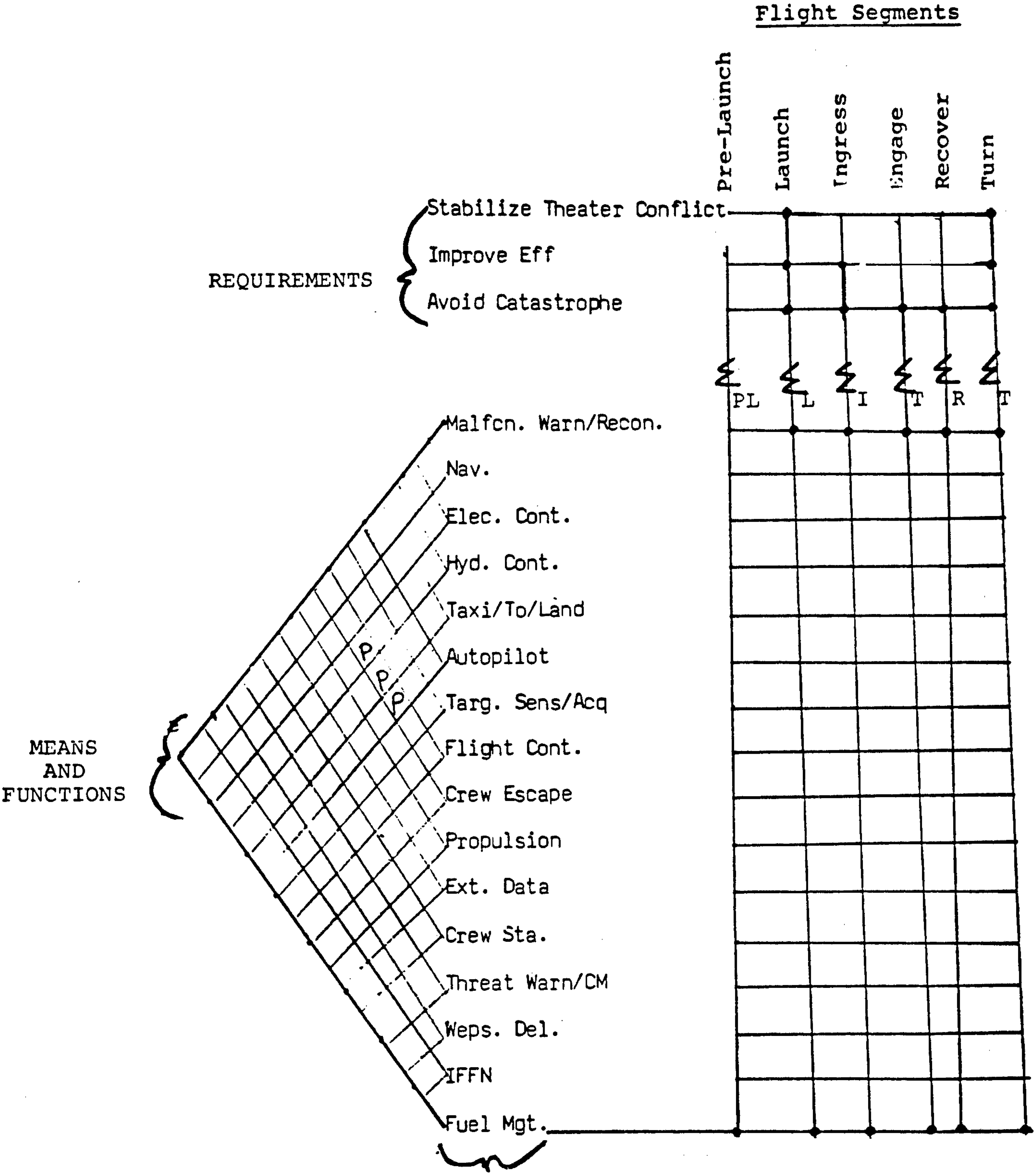


FIGURE 13. ANALYTIC METHODOLOGY

CREW STATION

MISSION SEGMENTS	PL	L	I	E	R	T
MISSION CONTORL FUNCTION	MISSION PLANNING	FLT PATH CONTROL COORDINATED	FLT PATH CONTROL TIME SPACED -SURVIVABLE	FLT PATH CONTROL SURVIVABLE TERMINAL -COORDINATED -MANEUVERING WEAPON DELVIERY	FLT PATH CONTROL SURVIVABLE	POST FLIGHT
MISSION OPERATIONAL FUNCTIONS	PLANNING METHODS BITE WEAPON LOADING & CHECKOUT	FLT CONTROL PROPULSION CONTROL NAVIGATION THREAT WARNING EXTERNAL DATA	FLT CONTROL PROPULSION CONTROL NAVIGATION TERRAIN F/A THREAT WARNING EXTERNAL DATA	FLT CONTROL PROPULSION CONTROL NAVIGATION TARGET ACQ/SEL. THREAT WARNING FIRE CONTROL IFFN	FLIGHT CONTROL PROPULSION CONTROL NAVIGATION THREAT WARNING IFFN	BITE DAMAGE ASSESS WEAPONS SEL.
MISSION SUPPORT FUNCTIONS	GROUND SUPPORT ELECT.	ELECTRICAL HYDRAULIC DATA DIRECT NETWORK WARNING SYSTEM CREW ESCAPE	ELECTRICAL HYDRAULIC DATA DIST NETWORK WARNING SYSTEM CREW ESCAPE	ELECTRICAL HYDRAULIC DATA DIST NETWORK WARNING SYSTEM CREW ESCAPE	ELECTRICAL HYDRAULIC DATA DIST NETWORK WARNING SYSTEM CREW ESCAPE	ELECT GROUND SUPPORT

1. INPUT SATURATION

2. MULTI-TASKING

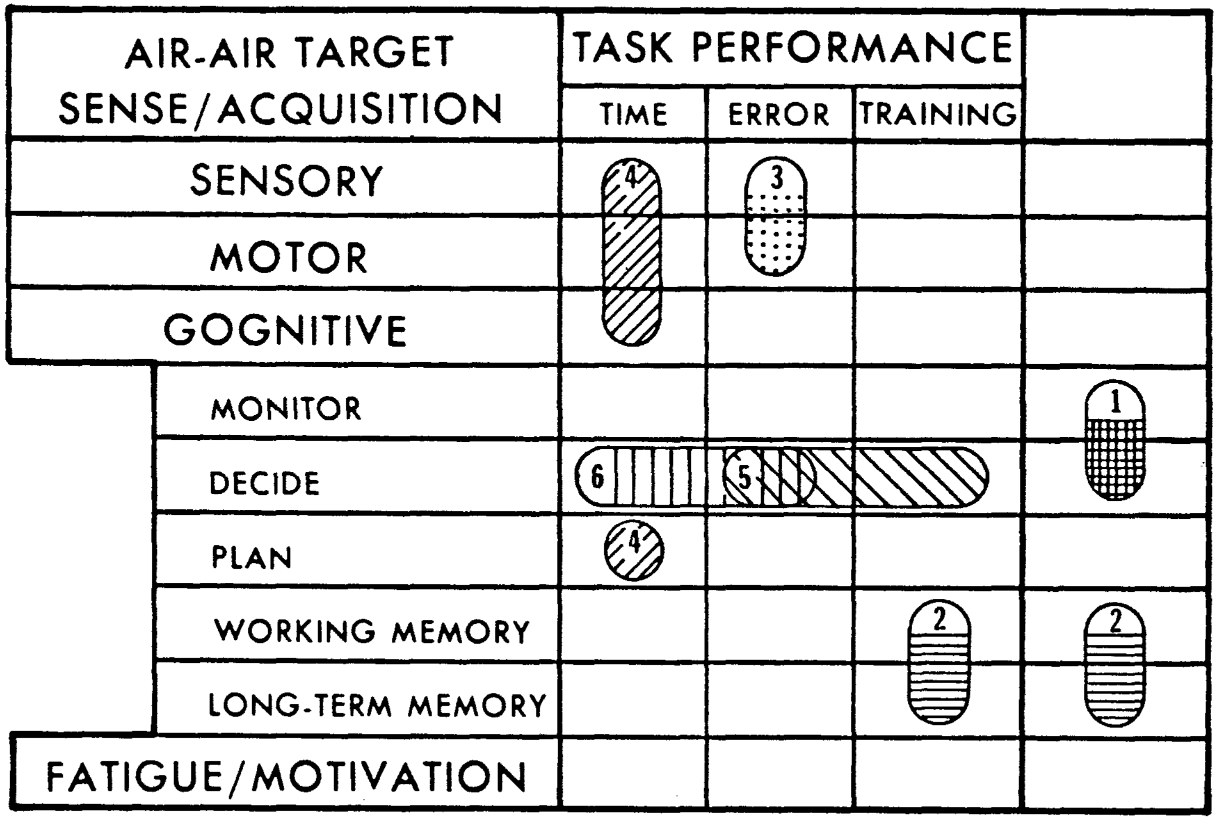
3. TIME-LINE COMPRESSION

4. PILOT BENDWIDTH LIMITATIONS

5. LOW-LEVEL OPERATIONS

FIGURE 15 SOURCES OF PILOT WORKLOAD

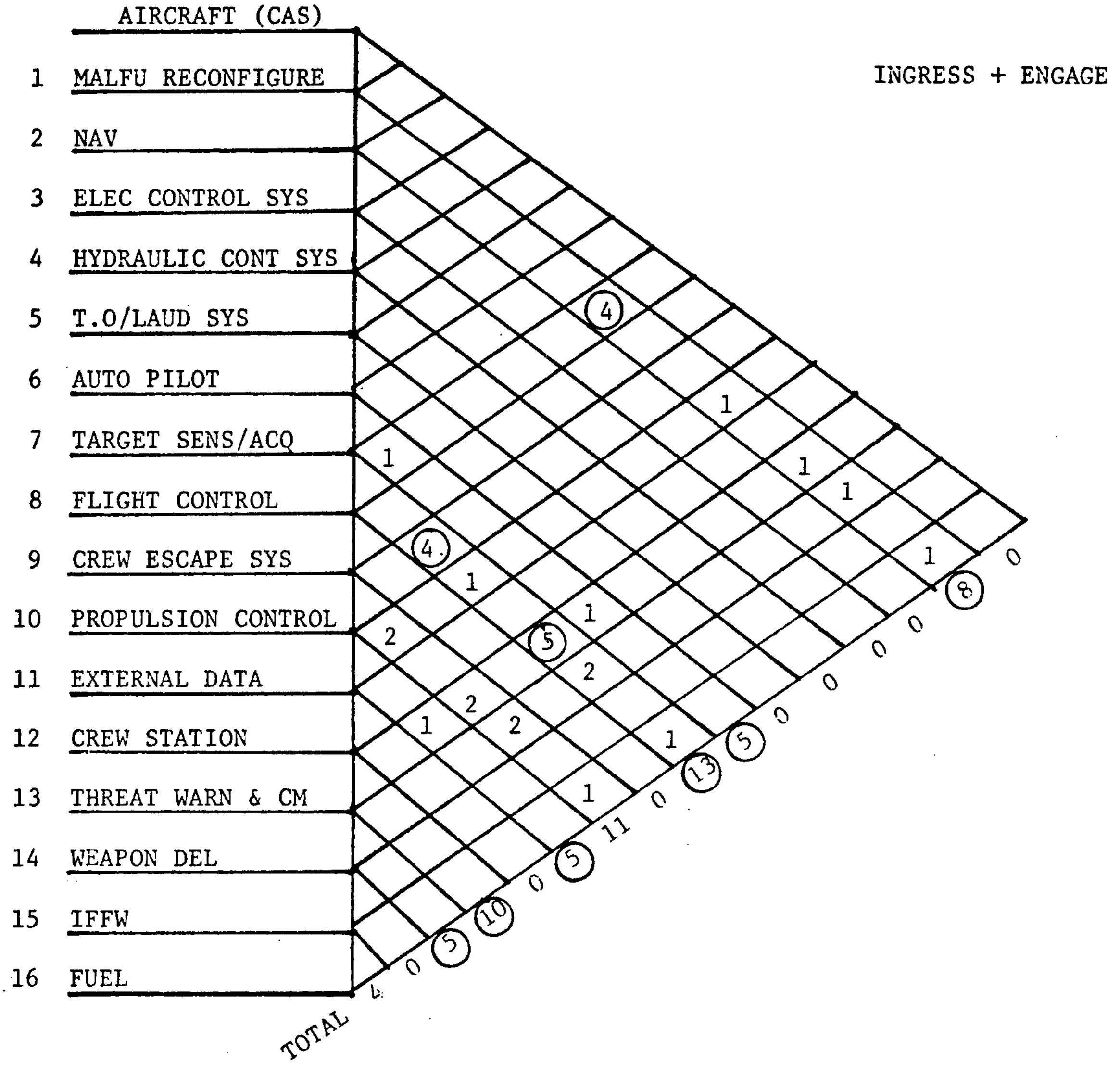
WORKLOAD GUIDELINES



GUIDELINES:

1. AUTOMATE ROUTINE FUNCTIONS
2. AUTOMATE MEMORIZATION TASKS
3. AUTOMATE PRECISION OPERATIONS
4. AUTOMATE SEQUENTIALLY TIMED OPERATIONS
5. DO NOT AUTOMATE WHERE PILOT JUDGEMENT REQUIRED
6. DO NOT AUTOMATE WHERE SITUATIONS VARIES AS A RESULT OF TACTICAL EVENTS

FIGURE 16



TOP 3 INTERACTIONS

① THREAT WARN & CM VERSUS FLIGHT CONTROL (5)

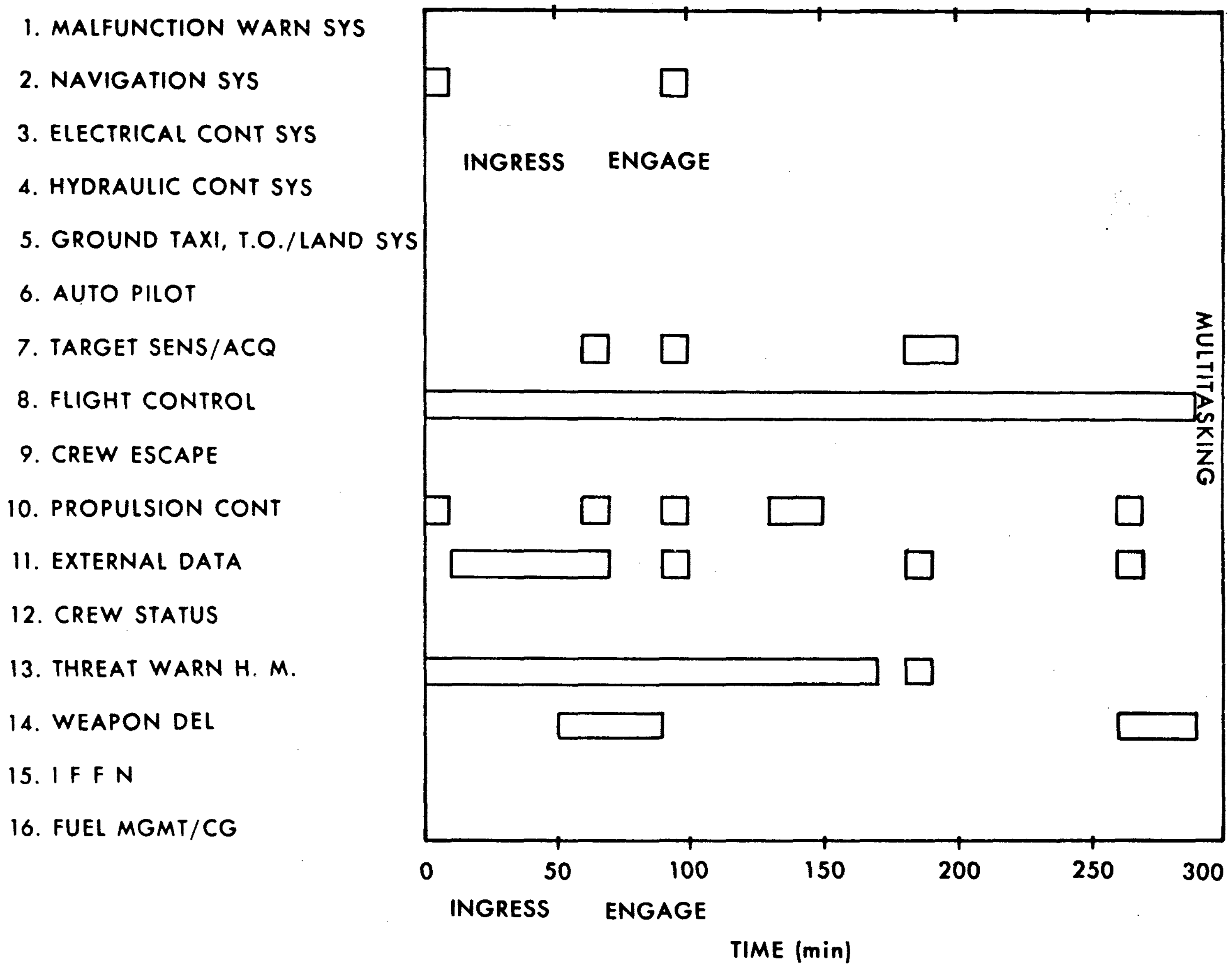
② PROPULSION CONTROL & FLIGHT CONTROL (4)

③ NAVIGATION VERSUS FLIGHT CONTROL (4)

TOP 5 SYSTEMS

1.	FLIGHT CONTROL	(13)
2.	THREAT WARNING & CM	(10)
3.	NAVIGATION	(8)
4.	TARGET SENS/ACQ	(5)
5.	EXTERNAL DATA	(5)
6.	WEAPON DEL	(5)

FIGURE 17 CREW INTERACTIONS


1. MALFUNCTION WARN SYS
2. NAVIGATION SYS
3. ELECTRICAL CONT SYS
4. HYDRAULIC CONT SYS
5. GROUND TAXI, T.O./LAND SYS
6. AUTO PILOT
7. TARGET SENS/ACQ
8. FLIGHT CONTROL
9. CREW ESCAPE
10. PROPULSION CONT
11. EXTERNAL DATA
12. CREW STATUS
13. THREAT WARN H. M.
14. WEAPON DEL
15. I F F N
16. FUEL MGMT/CG
INGRESS
ENGAGE
MULTITASKING
0
50
100
150
200
250
300
INGRESS
ENGAGE
TIME (min)

MIT PAPER	LITERATURE SURVEY
HIGHLY REPETITIVE FUNCTIONS	K CHARACTERISTIC (REGULATOR) FUNCTIONS
TOO DIFFICULT FUNCTIONS	HIGHER ORDER ON HIGH BANDWIDTH FUNCTIONS
TEDIOUS FUNCTIONS	REDUCE ORDER OF K/S^2 OR HIGHER ORDER FUNCTIONS
CRITICAL FAILURE FUNCTIONS	

TAC EXPERIENCE

- FAN ENGINE CONTROL
- WARNING SYSTEMS
- PRECISE FLIGHT CONTROL TASKS
- TARGET SENSOR INTEGRATION
- TARGET ACQUISITION/WEAPON DELIVERY
- THREAT WARNING/COUNTERMEASURES
- COMMUNICATION

FIGURE 19: SUMMARY - CANDIDATES CRITERIA AND AREAS FOR AUTOMATION

FIGURE 20 FLIGHT PATH CONTROL

Mission Phase

	Pre-Launch	Launch	Ingress
Present Approach	• Manual Mission Prep • Keyboard Data Entry	• Manual Control • Essential parameters on HUD	• Pilot Flys Computer Steering CMDS • Manual NAV Update • Manual Threat Avoidance
Automation Approach	• Auto mission prep station • Pre Fly Mission • Cassette Data Entry	• No Change	• Precise 4D NAV Coupled to Flight/Engine Control • Auto TF/TA • Auto Update of NAV and and Threat Data • Auto Threat Avoidance
Technology Needed for Automation Approach	• Data Base o Terrain o Threat o Targets • Compact Mass Memory	None	• Integraged Flight/Engine Control • Tightly Coupled TF/TA • Terrain and Threat Data Base • Compact Mass Memory • High Rate Data Network
Current Programs that Address Needed Technology	• DMA Data Base • CAMPS		• Interact (NASA) • LANTIRN • Blended TF/TA • GPS • JTIDS • PLSS • AETMS • TACTICAL Flight Management Program (AFFDL)
New Thrust or Further Emphasis Required	• DMA Data Base • Current Threat and Target Data Base • 100's Meg Bit Memory • Auto Mission Prep Station		• Compact Mass Memory • Threat Data Base • Blended TF/TA • Integrated Flight/Engine Control • High Speed Replacement for 1553

FIGURE 21 FLIGHT PATH CONTROL

Engage	Egress and Recovery	Turn-Around
• Pilot Steering to Accomplish Attack • Manual Threat Avoidance	• Same as Ingress and Launch	• Repeat Pre-Launch Functions • Replace defective LRU's
• Coupled Sensor/Flight/Engine Control • Auto Attack and Threat Avoidance • Task Tailored Control Laws	• Same as Ingress and Launch	• Reconfiguration/Fault Tolerant Systems • 100% Fault Isolation and Common Modules
• Coupled Fire/Flight/Engine Control	• Same as Ingress and Launch	• VHSIC • Distributed Functional Partitioning

• F15 IFFC
• AFTI-16
  • IFFC
  • AMAS
	• Same as Ingress and Launch	• DIGITAC III
		• Fault Tolerant Architectures (NASA)
		• Continuously Reconfigurable Flight Control System

• Fire/Flight/Engine Control Coupling	• Same as Ingress and Launch	• Distributed Functional Models

FIGURE 22 THREAT WARNING AND COUNTERMEASURES

	Pre-Launch	Launch	Ingress-Engagement
Present Approach	Refer to Flight Path Control ● Prebrief S to A threats		● Verbal In-Flight Threat Info ● SF-111 ● On-board TWS ● Manual Threat Avoidance ● Auto and Manual CM
Automation Approach	Refer to Flight Path Control	Refer to Flight Path Control	● Threat Data Base ● Real-time threat update ● Auto response to new threat
Technology Needed for Automation Approach	Refer to Flight Path Control		● Stored Threat Data Base ● Mass Memory ● Data Link ● Fire/Flight/TWS Integration
Current Programs that Address Needed Technology	Refer to Flight Path Control		● PLSS ● JTIDS ● Purple Haze ● ASPJ
New Thrusts Required-or-Added Emphasis	Refer to Flight Path Control		● Continuous Threat Data Base ● Mass Memory ● PLSS ● Data Link ● Fire/Flight/TWS Integration

FIGURE 23 THREAT WARNING AND COUNTERMEASURES

Egress & Recovery	Recovery	Turn-Around
● Repeat of Ingress and Launch Functions --Plus-- ● Egress Procedures	o Same as Launch Function	o Refer to Weapon Delivery A-A Function
● IFF with friendly troops		
● New IFFN		
● New IFFN		

FIGURE 24 TARGET SENSING AND ACQUISITION

	Mission Phase			
	Pre-Launch	Launch	Ingress	Engage A/A
Present Approach				• Mostly Manual • Some Semi-Automatic
Automation Approach				• Automatic detection, acquisition, identification, and prioritization of targets
Technology Needed for Automation Approach				• Beyond Visual Range I.D. • Multi-sensor correlation • External data correlation • Multi-target acquistion
Current Programs that Address Needed Technology				• JTIDS • IFFN Fusion • Jet engine modulation I.D. in F15
New Thrusts Required-or-Added Emphasis				• Beyond Visual Range I.D. • Multi-sensor correlation • Light weight helmet sight and display • IR Search & Track

FIGURE 25 TARGET SENSING AND ACQUISITION

Mission Phase

Engage A/G	Egress & Recovery	Turn-Around
• Manual		
• Automatic detection acquisition, identification, and prioritization of targets		
• Automatic target detection, calssification and identification • High resolution sensors		
• ERIM Ultra-high resolution radar • Multiple Source Integration • JTIDS • Covert Strike • PAVE MOVER • LANTIRN • PLSS • Automatic SAR Target Classification • Correlated Sensor Data Display		
• Auto target pattern recognizer • Multisensor correlation • High resolution sensors • Light weight helmet sight and display		

FIGURE 26 WEAPON DELIVERY (Air-Air)

	Mission Phase		
	Pre-Launch	Launch	Ingress
Present Approach	● Manual Mission Prep ● Keyboard Data Entry	● Verbal Comm link for target update	● Verbal comm for target assignment ● Refer to Flight Path Control Function
Automation Approach	● Auto Mission Prep Station ● Pre Fly Mission ● Cassett Data Entry	● Auto target update	● Real time display for targets designated by each attack airplane ● Refer to Flight Path Control Function
Technology Needed for Automation Approach	● DMA Data Base ● Current Threat and Target Data Base ● 100's Meg Bit Memory	● Data Link	● Data Link
Current Programs that Address Automation Needs	● DMA Data Base ● Camps	● JTIDS	● JTIDS

New Thrusts Required or Added Emphasis	● Data Base
	  ● Terrain
	  ● Threat
	  ● Targets
	● Compact Mass Memory
	● Auto Mission Prep Station		

FIGURE 27 WEAPON DELIVERY (AIR-AIR)

Engage	Turn-Around	Egress and Recovery
● Visual and radar detect, visual ID, auto launch zone computation, pilot null steering, manual weapon release	● Repeat Pre-Launch Functions ● Replace defective LRU's	● Refer to Flight Path Control Ingress Function
● BVR auto detect and ID ● Auto threat prioritization, steering and weapon release	● Reconfiguration/fault tolerant systems ● 100% fault isolation and common modules	
● BVR ID ● Priority Algorithms ● IFFC ● Integ Engine/Flight Controls	● VHSIC ● Distributed Functional partitioning	
● Multiple Source Integration ● JTIDS ● TAACS ● MISVAL ● AFTI-16 ● F-15 IFFC ● Interact (NASA)	● DIGITAC III ● Fault Tolerant Architecture (NASA) ● Continuously Reconfigurable Flight Control System	
● AA IFFN ● Priority Algorithms ● Integ Engine/Flight Controls	● Distributed functional modules	

FIGURE 28 WEAPON DELIVERY (AIR TO GROUND)

	MISSION PHASE		
	Pre-Launch	Launch	Ingress
Present Approach	● Manual Mission Prep ● Keyboard data entry	● Verbal Comm Link for Target Update	● Target assignment via verbal coordination ● Refer to Flight Path Control Functions
Automation Approach	● Auto Mission Prep Station ● Pre fly mission ● Cassette data entry	● Auto target update	● Real time display of target location ● Refer to Flight Path Control Functions
Technology Needed for Automation Approach	● Data Base ● Terrain ● Threat ● Targets ● Compact Mass Memory	● Data Link	● Data Link ● Tactical Flight Management
Current Programs that Addresses Needed Technology	● DMA Data Base ● CAMPS	● JTIDS	● JTIDS ● Tactical Flight Management
New Thrusts or Added Emphasis	● DMA Data Base ● Current threat and target data base ● 100's Meg Bit Memory		

FIGURE 29 WEAPON DELIVERY (AIR TO GROUND)

Engage	Egress and Recovery	Turn-Around
● Pilot Visual Detect and Manual Steering (CCIP) ● Pilot Visual Detect and Auto Delivery (Dive Toss) ● Pilot ID & Designate (Radar) with Manual Steering (Auto Rel) ● Pilot Visual ID & Designate (Laser, E.O.) ● Manual Release ● Pilot Visual ID (Guns) Manaul Steering-Pipper	Refer to Flight Path Control Ingress Function	Refer to Weapon Delivery (Air to Air)
● Auto detect, ID & classification ● IFFC/AMAS ● Auto computer in flight weapon fusing		
● Hi Res. Sensors, Auto pattern Recog ● Auto-Correlation Tech ● IFFC/AMAS ● Auto Fire Control Fusing		
● LANTIRN ● PAVE MOVER ● Covert Strike ● JTIDS ● F-16 IFFC ● AFTI-16 AMAS ● SAIF		
● Hi Resolution AIG Sensor ● Pattern Reconfiguration ● Auto Correlation Tech		

STRUCTURAL INTEGRATION/AUTOMATION

APPROACH REQUIREMENTS ESSENTIAL FOR

AUTOMATED COMBAT AIRCRAFT

FIGURE 30

<u>EXAMPLES</u>:

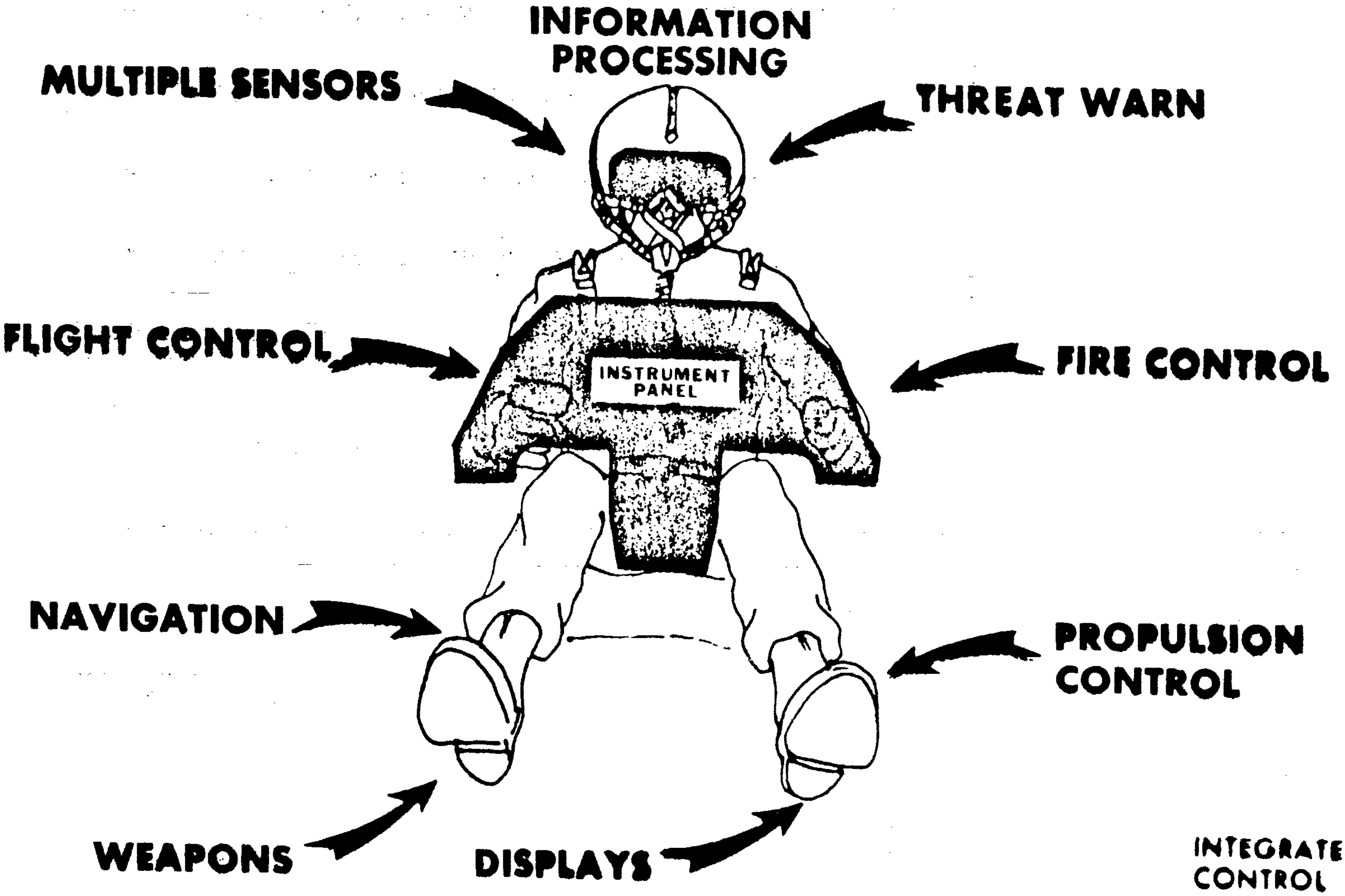
- FIRE CONTROL
- PROPULSION CONTROL
- FLIGHT CONTROL
- SENSORS
- WEAPONS
- THREAT WARM

<u>PROBABLE REASONS</u>:

- SPECIFIC FUNCTIONAL BOTTLENECKS ATTACKED
- LOGICAL EXTENSION OF EXISTING DESIGNS
- ORGANIZATIONAL ALIGNMENTS
- LIMITED CROSS-TECHNOLOGY INTEGRATION KNOWLEDGE

FIGURE 31: HIGH LEVELS OF AUTOMATION INCORPORATED BUT LIMITED TO FUNCTIONAL AREAS

PILOT IS "AUTOMATION CORE"



INFORMATION PROCESSING

MULTIPLE SENSORS

THREAT WARN

FLIGHT CONTROL

INSTRUMENT PANEL

FIRE CONTROL

NAVIGATION

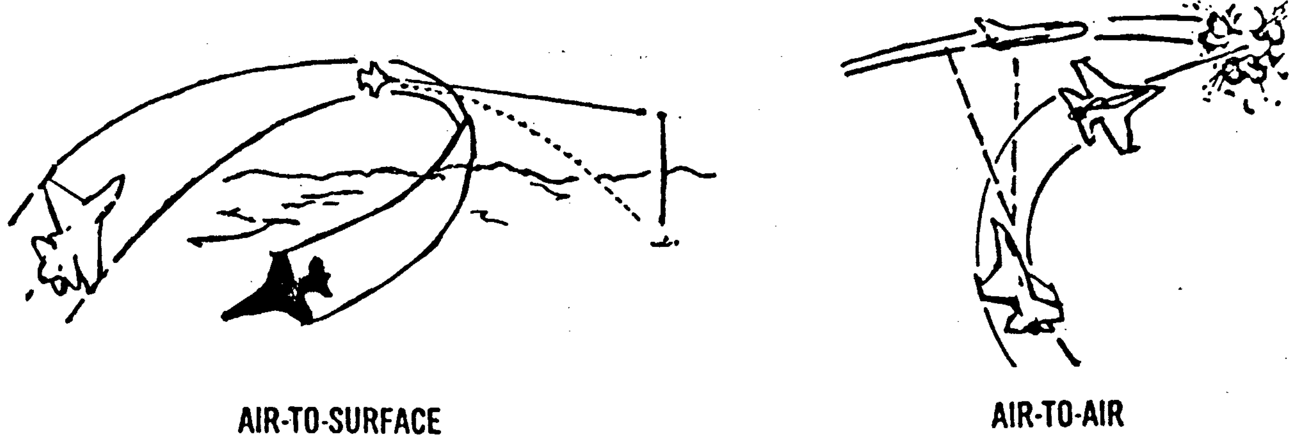
PROPULSION CONTROL

WEAPONS

DISPLAYS

INTEGRATED
CONTROL
TECHNOLOGIES

FLIGHT TRAJECTORY AND ATTITUDE FUNCTION (WITH TIME) IS LOGICAL CORE CANDIDATE



- COMMON FUNCTION
- CENTRAL TO AIRPLANE ACTIVITY
- ENCOMPASSES SEVERAL FUNCTIONAL AREAS
  - FIRE CONTROL
  - NAVIGATION
  - FLIGHT CONTROL
  - PROPULSION

FIGURE 33: SOLUTION IS TO REALLOCATE FUNCTIONS BY IMPLEMENTING A DIFFERENT "AUTOMATION CORE"

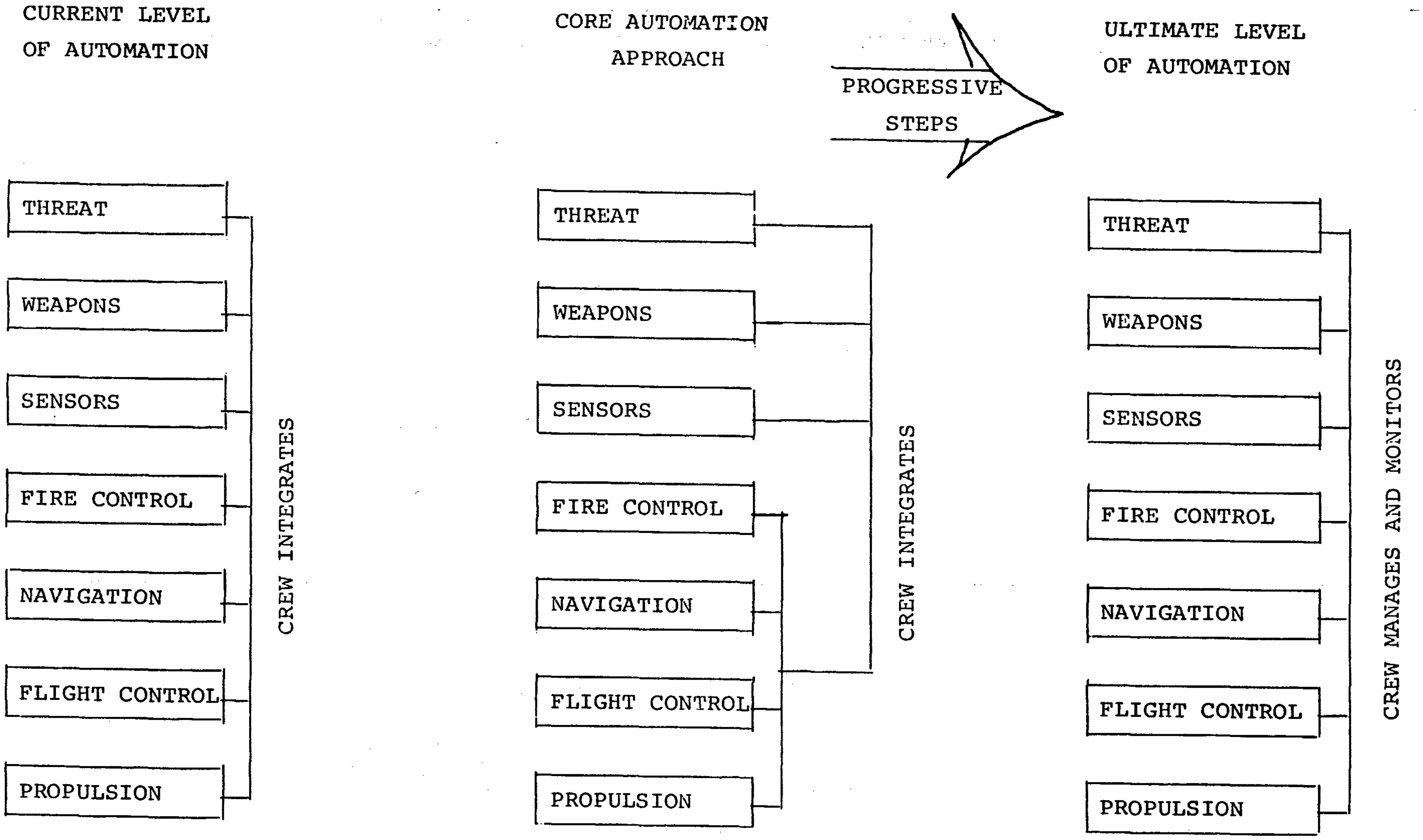
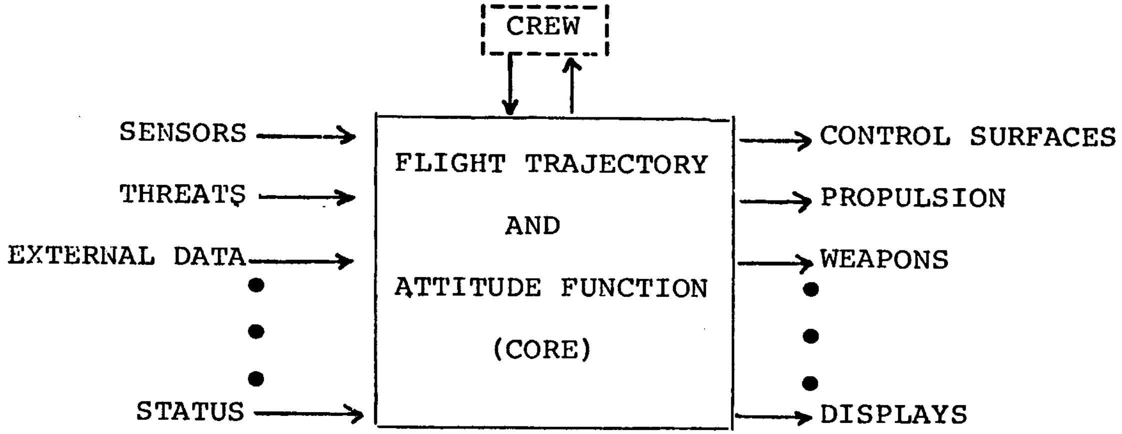


FIGURE 34: LOGICAL AUTOMATION PROGRESSION



- UNIVERSAL "INTERFACE" CORE
- BASIS FOR INPUT AND OUTPUT SPECIFICATIONS
  - COMMON EQUIPMENTS
  - COMMON SOFTWARE ELEMENTS
- LOGICAL GROWTH TO HIGHER AUTOMATION LEVELS AS INPUT HARDWARE IMPROVES

FIGURE 35: CORE APPROACH AND DEVELOPMENT ESSENTIAL FOR CONTINUED AUTOMATION

- DISTRIBUTED OR CONCENTRATED REDUNDANT PROCESSORS
- FIRMWARE OR HARDWARE
- SINGLE BUS, MULTIPLE PARALLEL BUS, HIERARCHY OF BUSES OR BUS NETWORK
- STANDARDS APPLICABLE TO MULTIPLE FUNCTIONAL AREAS

FIGURE 36: "AUTOMATION CORE" IS A SYSTEM STRUCTURE WITH LOGICAL PARTITIONING AND INTERFACE STANDARDS

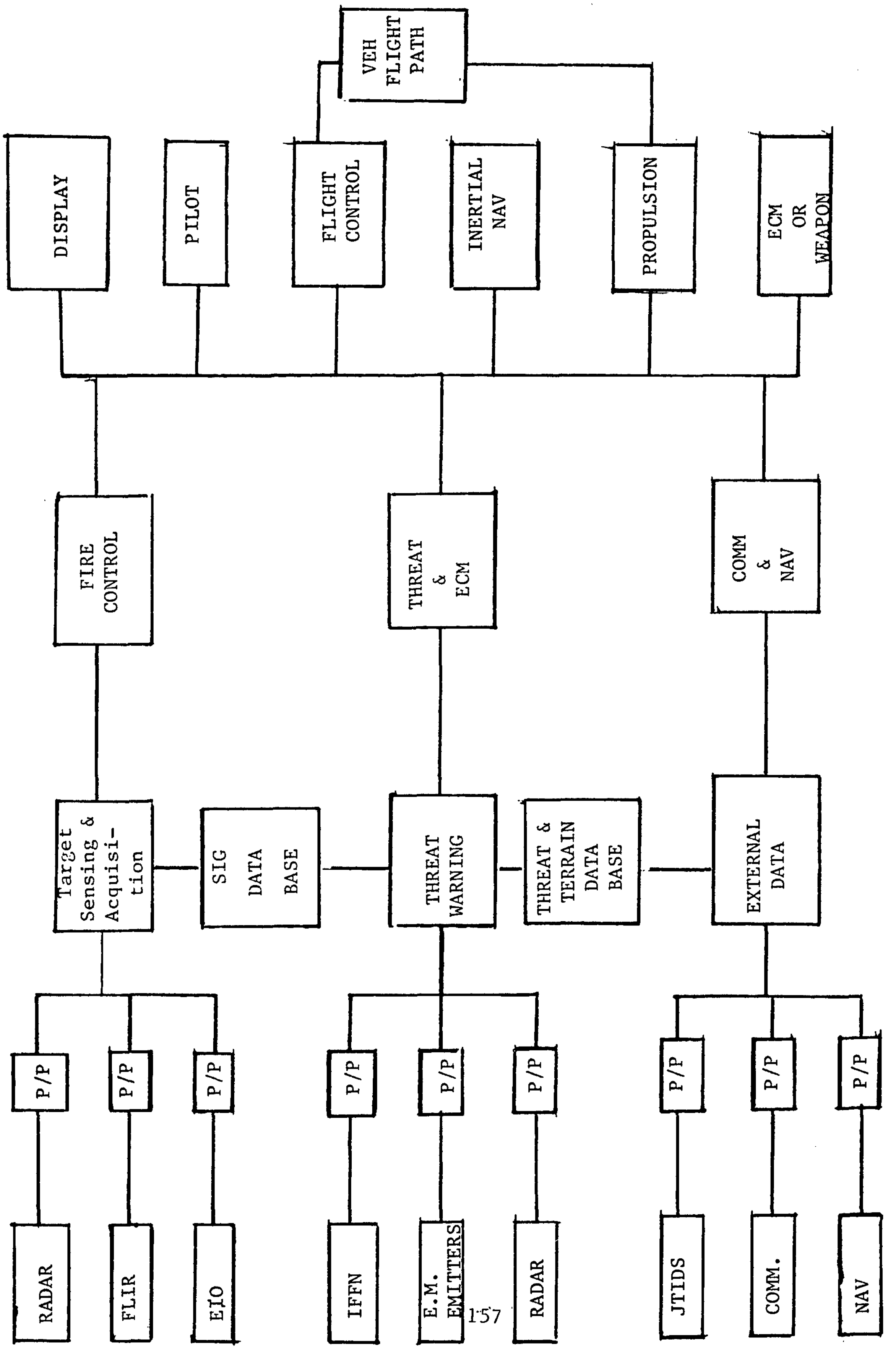


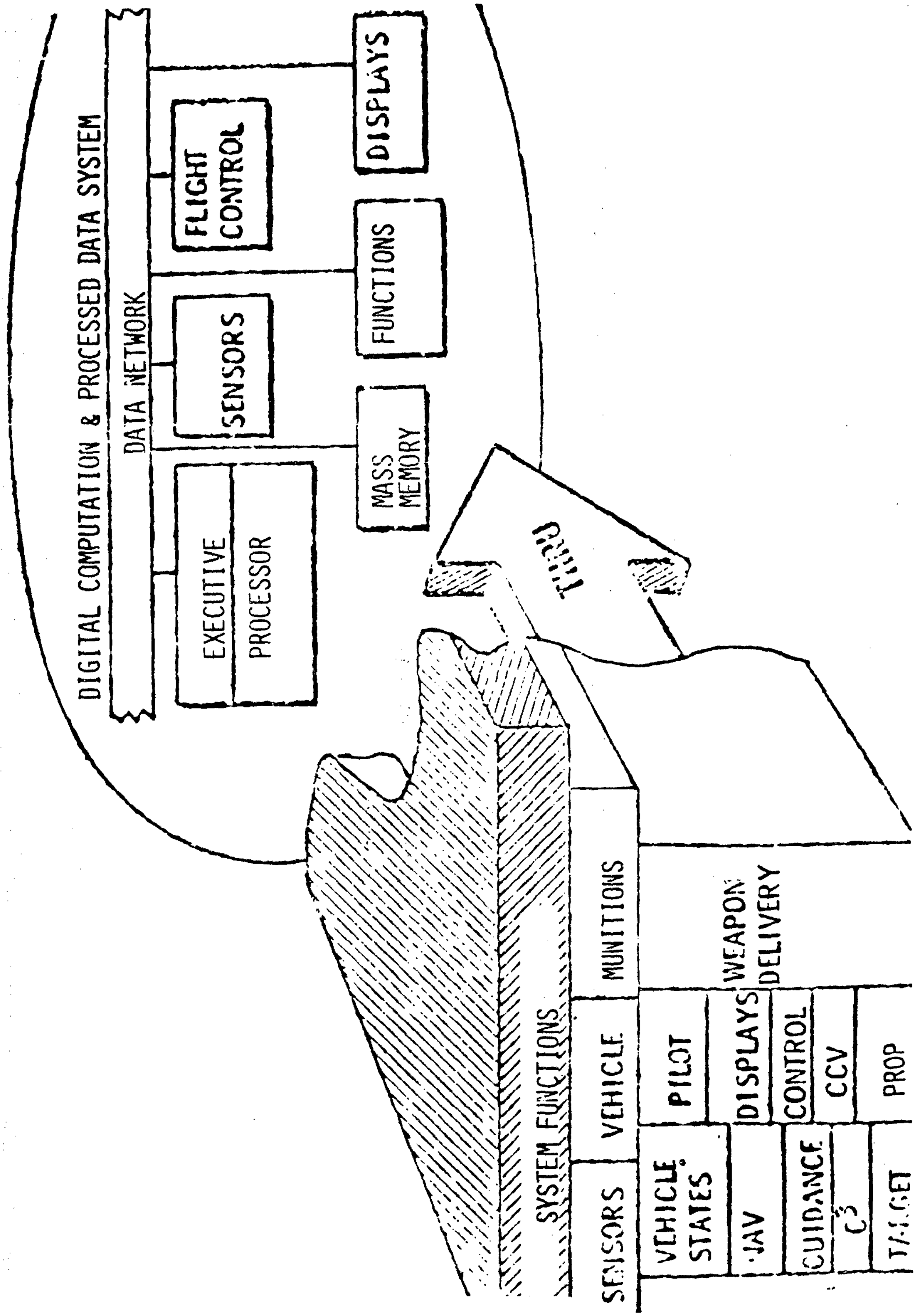
FIGURE 37: FUNCTIONAL INFORMATION FLOW

- INCREASED EMPHASIS ON PRECISION FLIGHT PATH CONTROL
- INCREASED EMPHASIS ON FLIGHT CONTROL - PROPULSION CONTROL - NAVIGATION AND FIRE CONTROL INTERACTION AND INTEGRATION
- INCREASED DEPENDENCE UPON A COMMON DATA DISTRIBUTION NETWORK

FIGURE 38: FUTURE TRENDS

DEVELOPMENT OF A CORE ARCHITECTURE WITH THE CAPACITY - INTEGRITY AND FLEXIBILITY TO SUPPORT THE TRENDS.

FIGURE 39: CHALLENGE


DIGITAL COMPUTATION & PROCESSED DATA SYSTEM
DATA NETWORK
EXECUTIVE
PROCESSOR
SENSORS
FLIGHT CONTROL
MASS MEMORY
FUNCTIONS
DISPLAYS
THRU
SYSTEM FUNCTIONS
SENSORS
VEHICLE
MUNITIONS
VEHICLE STATES
NAV
GUIDANCE
C³
TARGET
PILOT
DISPLAYS
CONTROL
CCV
PROP
WEAPON DELIVERY

REFERENCES

1. D. McRuer and D. H. Weir, Theory of Manual Vehicular Control; Erogomics, 1969, Vol 12, No. 4, 599633.

2. M. A. Ostgaard, "Active Control Technology"; VKI Lecture, December 1978.

3. Daniel S. Berlinger, M.D., Lt Col., "Medical Aspects of Helicopter Evacuation and Rescue Operations"; U.S. Army Agency for Aviation Safety, AGARD Proceedings, AGARD-CP-255, 1-5 May 1978.

4. R. Beyer, "Flight Performance and Pilot Workload in Helicopter Flight Under Simulated IMC Employing a Forward Looking Sensor"; AGARD Proceedings AGARD-CP-240, 17-20 October 1977.

5. G. A. Barnes, "Guidance and Control for Low Level Offensive Aircraft---A Royal Air Force View"; AGARD Proceedings AGARD-CP-240, 17-20 October 1977.

6. C. W. Brinkley, P.S. Sharp, and R. Abrams, "B-1 Terrain-Following Development"; AGARD Proceedings AGARD-CP-240, 17-20 October 1977.

7. J. F. Moynes and J. T. Gallagher, "Flight Control System Design for Ride Qualities of Highly Maneuverable Fighter Aircraft"; AGARD Proceedings AGARD-CP-240, 17-20 October 1977.

8. J. G. Jones and D. E. Fry, "Aircraft Ride-Bumpiness and the Design of Ride-Smoothing Systems"; AGARD Proceedings AGARD-CP-240, 17-20 October 1977.

COCKPIT DESIGN FOR THE FUTURE
AND
CHALLENGES TO WORKLOAD MEASUREMENT

BY

LARRY C. BUTTERBAUGH
FLIGHT DYNAMICS LABORATORY
AIR FORCE WRIGHT AERONAUTICAL LABORATORIES
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

INTRODUCTION

The implementation of current and developing technologies in the area of airborne subsystem avionics, and control and display hardware, will have a dramatic impact on future cockpit designs, crew workloads, and design and evaluation techniques. However, before speculating on the design of future cockpits and the workload problems they will pose, let's briefly overview the last 50 years of cockpit technology and crew workloads.

The SPAD (Fig. 1), a World War I fighter, incorporated early display technology for subsystem monitoring which consisted of in-line gauges. However, with few subsystems to monitor, and skilled flyers using "seat-of-the-pants" and engine noise/vibration information as indicators of system operation, workloads associated with these tasks were low. But, these low workloads were probably replaced by other workloads associated with the actual flying task. By the time the P-26 (Fig. 2) was flying in the mid 1930s, flight instrumentation such as airspeed and altitude were being displayed along with subsystem information. However, the amount of information to monitor was still fairly small.

Moving ahead to World War II, the P-51 (Fig. 3) had quite a complement of subsystem controls and displays in addition to attitude and navigation information. And, of noteworthy importance with regard to the pilot's workload, was the increased speed capability of these aircraft coupled with more tasks to accomplish. In other words, the time required-time available ratio (a long-time standard for workload assessment) was on the rise. An additional 20-year jump brings us to the mid 1960s, and the F-111 (Fig. 4), which requires considerable monitoring and inputs.

Now in retrospect (Fig. 5), we know that more sophisticated missions and the inclusion of advanced subsystems to assure completion of these missions were straining not only the physical space of the cockpit, but also placing extremely high task loadings on the crew. Fortunately, however, display and avionics technonogy has not been standing idle. Digital avionics and multi-purpose displays have found a place in the cockpit, as illustrated by the Navy's F-14 (Fig. 6). The incorporation of such displays allows more efficient and effective display of information, by displaying only what's needed, when it's needed. As a result, workloads are able to be held down tc a tolerable level, simultaneous to achieving desired system and mission performance. The continued development of display technologies to go with avionics developments, therefore, make it very probable that the next generation fighter cockpit may resemble that of the Navy's F-18 (Fig. 7).

This introduction has quickly arrived at the state-of-the-art regarding cockpit technologies. But, more importantly, we've witnessed a workload transformation. Easily measured and quantified workloads associated with task accomplishment has been replaced by the more subtle and difficult to measure monitoring and decision-making workloads. And, as will be pointed out, the cockpit design problem will not be getting any easier, especially from the standpoint of advanced controls and displays. What is hoped for is that development of workload assessment techniques will be responsive to the designer's needs.

ADVANCED COCKPITS

The technologies being developed in the laboratories today will impact future cockpit designs in many ways. Some of these will include the functional allocation of tasks between the man and the machine, the modality of subsystems interface, advanced display hardware, and display formats and symbology.

Function Allocation

The autopilot of current aircraft is one example of computer-aiding the flying task by functionally allocating selected operations to be performed automatically. This allocation has long been shown to provide better tracking performance while at the same time reducing, or at least altering, the workload of the pilot. However, artificial intelegence algorithms (Fig. 8) may one day be put to use in the cockpit to assist the pilot in decision-making situations.

Subsystem Interface

While manual input and visually displayed output have been the longstanding tradition of cockpit design, single alphanumeric keyboards that replace the knobs and switches of several seperate subsystems are state-of-the-art. Voice technology (Fig. 9), however, is rapidly developing into a workable subsystem interface modality. Studies have already been conducted (Fig. 10) to identify candidate uses for speech recognition, as well as speech generation (Fig. 11) of advisories, warnings, and status reports.

Advanced Displays

Underlying the potential application of computer generated imagery and electronically projected images from radar, tv, and other sensors, is the development of advanced display hardware. The CRT (Fig. 12) has become a feasible display device and has been implemented in numerous cockpits. Flat-panel devices, such as the LED matrix display, are being developed as a potential substitute, in certain cases, and have several advantages over the CRT in weight, power and depth requirements.

Display Formats/Symbology

The most significant impact of these display technologies is the versatality of the way the information is presented. The first generation EADIs (Fig. 13) showed no originality and simply repeated on a video screen a two-dimensional picture of the electromechanical ADI. There is, however, significant work underway to study innovative pictorial formats and symbology. The following illustrate some of these concepts:

Fig 14 - A VSD format combining attitude, navigation, and tactical situation data;

Fig 15 - An HSD format showing navigation features and flight plan data;

Fig 16 - Another HSD format from a higher altitude and showing the display with and without clutter;

Fig 17 - Using computer-generated imagery, a head-up view can be projected heads-down to aid in transition;

Fig 18 - A HSD format with range limitations overlayed, to give status information;

Fig 19 - A MPD format with procedural information;

Fig 20 - A MPD format with advisory and procedural information.

DESIGN TECHNIQUES

One can expect that with the implementation of any or all of these cockpit technologies, the job of the crew station designer will not get any easier. The laboratory-based design process will have to continue to be mission oriented and involve the user in the definition and evaluation of the design. Further, because of the complexity of the systems and the costs to simulate them, confident design decisions will be necessary early in the process, requiring accurate and reliable analytic evaluation tools. Also, the techniques applied in the laboratory should be flight-test validated, so that laboratory results will be confidently accepted as predictive of flight-test and operational performance.

Mission-Based Design

Subsystem and control/display definition based on mission objectives and requirements is imperative (Fig. 21). The multiple-mission concept implies varying control and display requirements and workloads. To ensure the cockpit design satisfies all mission requirements, it is necessary to evaluate the design with regard to the total mission.

User Involvement

The involvement of pilots in the design process reaches beyond their role in man-in-the-loop simulation and effectively begins with design definition (Fig. 22). Pilots carry around expertise that can aid the designer in understanding what is required, or desired, from the operational vantage point. Pilots can translate mission requirements into control and display requirements. They can evaluate candidate designs and provide the designer with valuable evaluation data.

Quantitative Techniques (Fig. 13)

With the costs of simulation and flight-test becoming more and more expensive, the value of quantitative design evaluation techniques will become apparent in the design of future cockpits. Design alternatives will have to be evaluated prior to simulation with regard to a variety of parameters, preventing valuable simulation time from being used for an inferior design. Some of these analytic techniques exist, and others remain to be developed.

Flight-Validated Techniques (Fig. 24)

Related to the need for quantitative techniques applicable early in the design process, is the need for laboratory techniques to be cross-validated with flight-test and operational data, in order to ensure their sensitivity and accuracy. The more reliable, accurate, and proven the techniques used in early design, the more confidence the designer will have in the data and his decisions.

WORKLOAD MEASUREMENT CHALLENGES

Subjective Techniques (Fig. 25)

With complex missions, workloads that are difficult to measure, and the ability of pilots to integrate their experiences into a determination of workload, subjective techniques are as important in the assessment of workload as any quantitative method. Subjective methods will continue to be valuable tools, especially because of the cockpit trend towards more monitoring and decision-making tasks for the crew (the kinds of workloads, unfortunately, which are hard to assess). Presently there is need of subjective techniques which produce ratio-scale data. It will take this type data for meaningful comparisons among alternative designs to be possible. Fortunately, recent work has been completed and current work is underway which addresses this area of subjective workload assessment.

Measures and Criteria

The measurement of workload, with physiological and/or psychological methods, will also find an important role in advanced cockpits. One unique opportunity (Fig. 26) is the use of sensitive, validated measures in an adaptive crew system concept. In this application, pertinent physiological and/or task performance parameters would be monitored, preferably unobtrusively, and onboard algorithms would determine what flight control task, or other tasks, should be transferred to the machine, or onboard computers, for accomplishment. Such an application, however, depends on the establishment of criteria levels for the parameters monitored. This will not be easy, since this has been a continuing problem since the inception of the concept of workload and the desire to evaluate it.

SUMMARY

In conclusion, workload assessment appears to have a major role to play both in the design of future cockpits, and in its application as a component of the crew system. The challenge to the development of workload measures and predictors is the need for sensitive, diagnostic techniques in the design process, and quantifiable, criterional measures for use in the flight-test and operational environments.

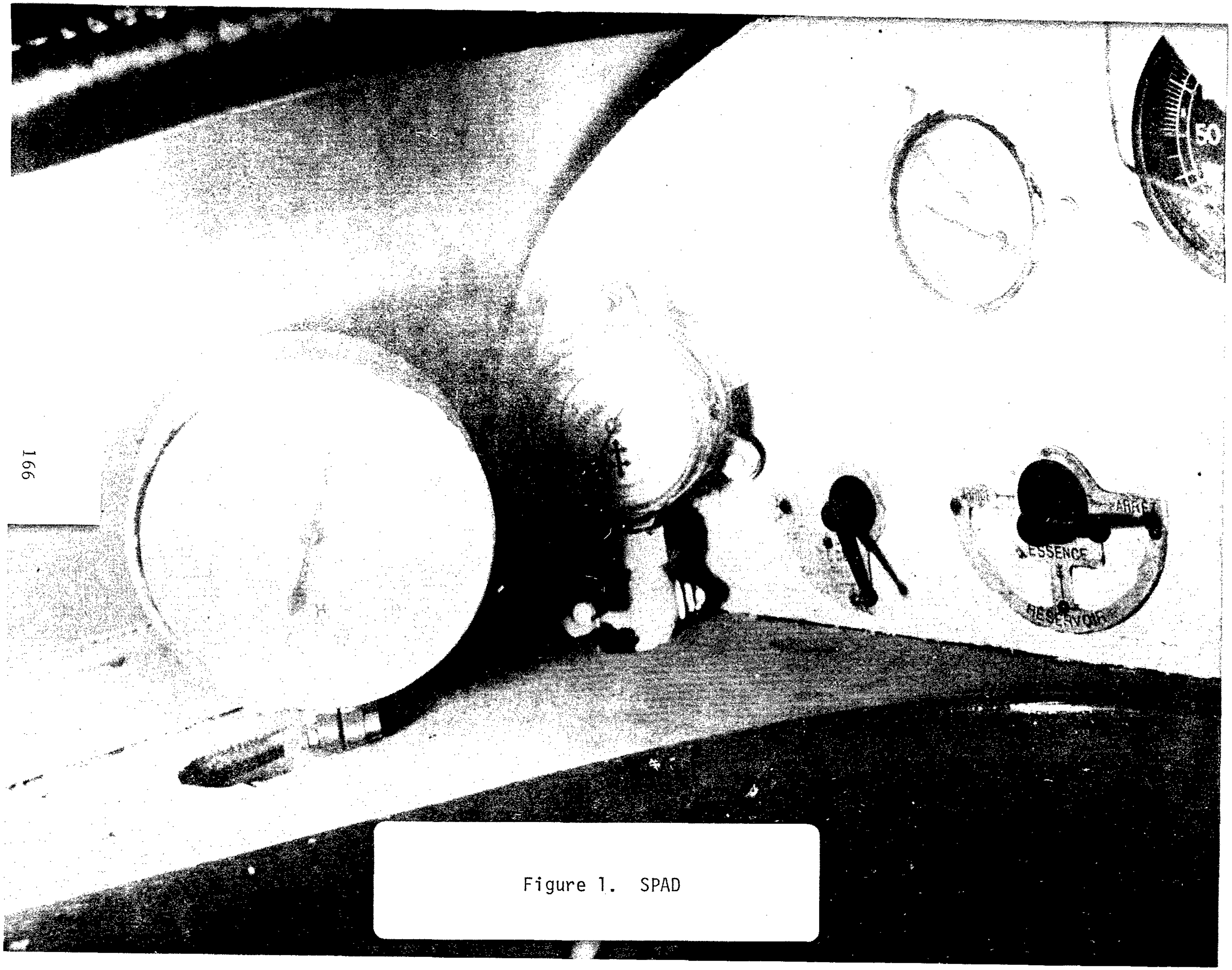


Figure 1. SPAD

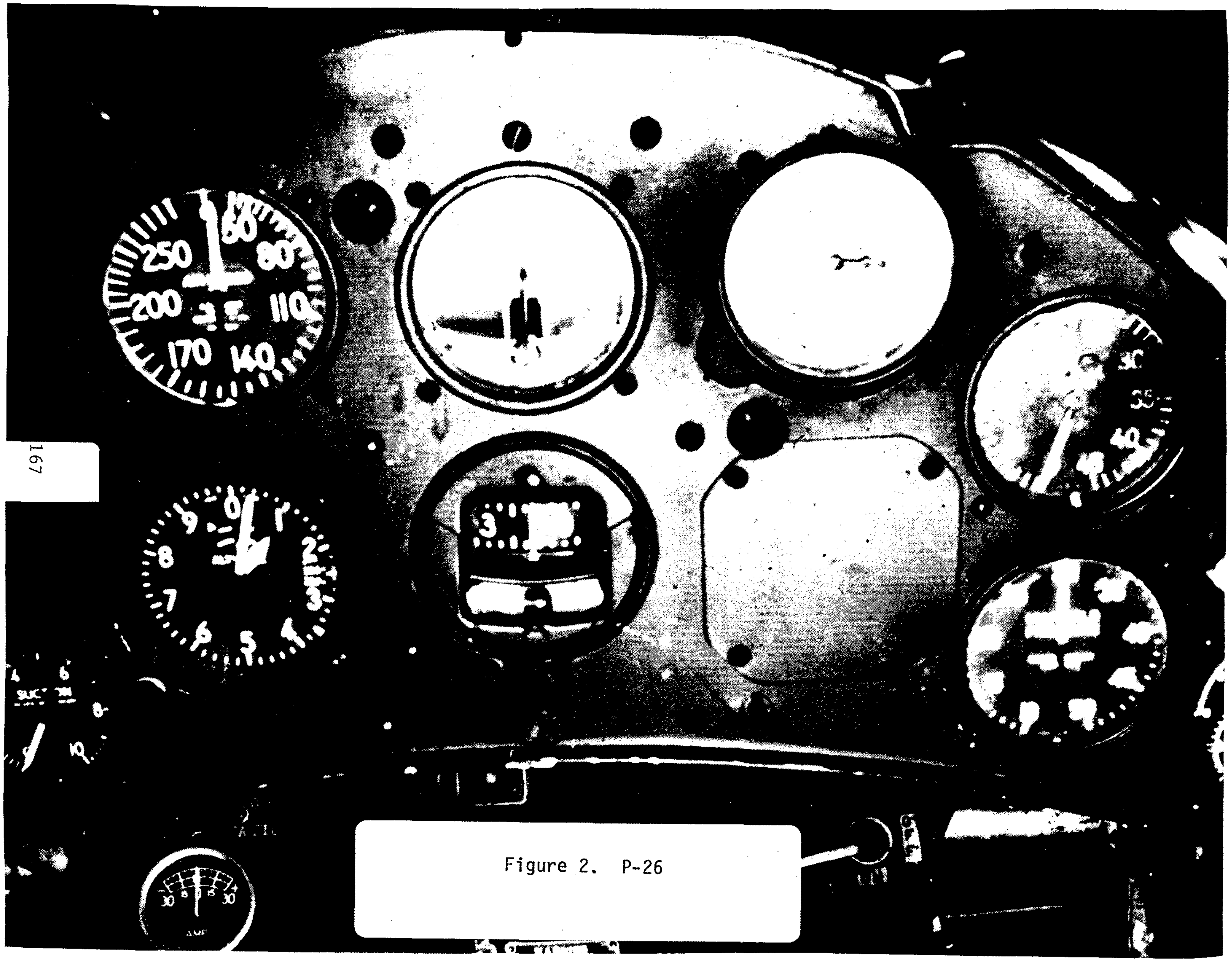


Figure 2. P-26

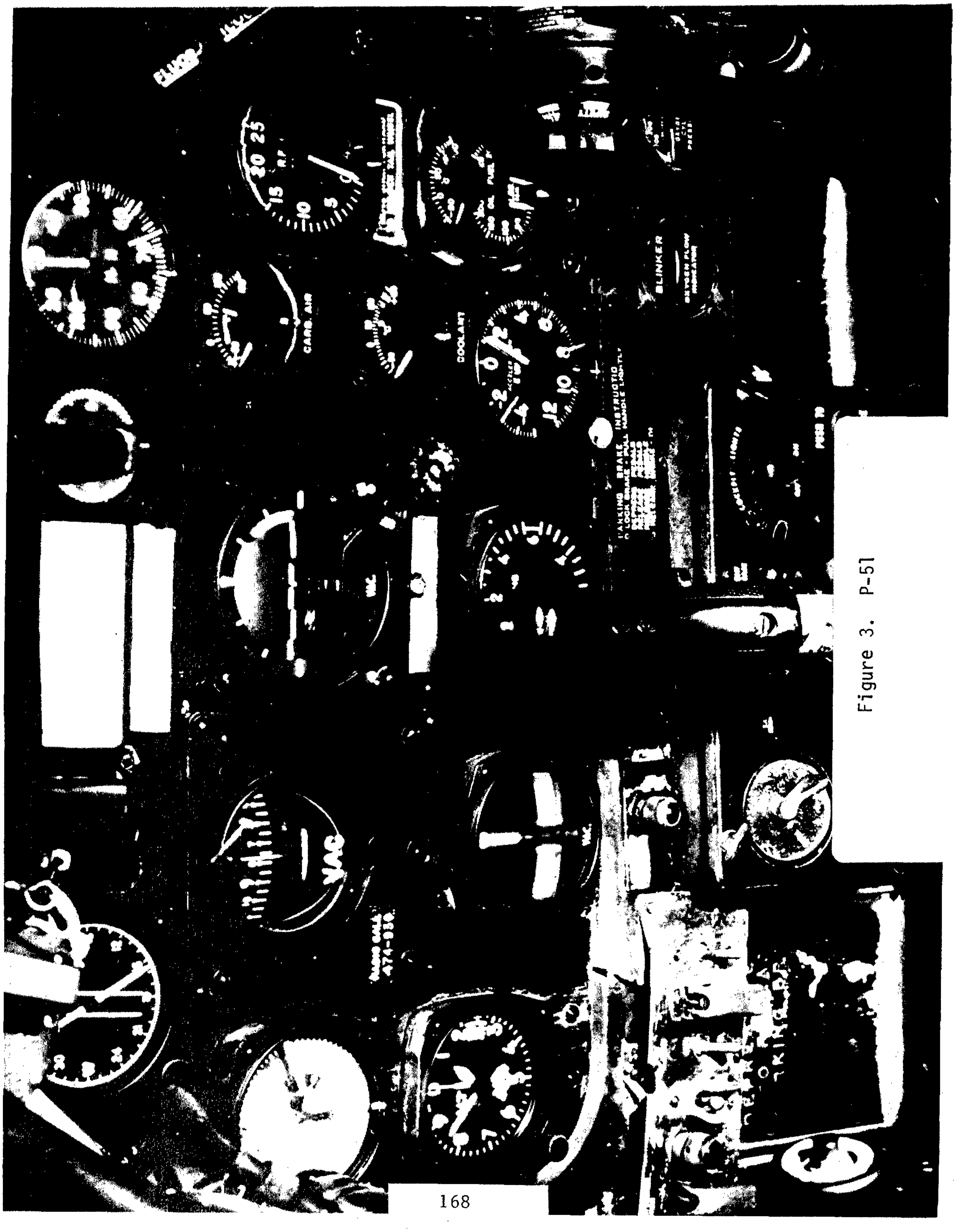

CARB. AIR
COOLANT
BLINKER
OXYGEN FLOW INDICATOR
PARKING BRAKE INSTRUCTIO
RADIO CALL 474-936

Figure 3. P-51

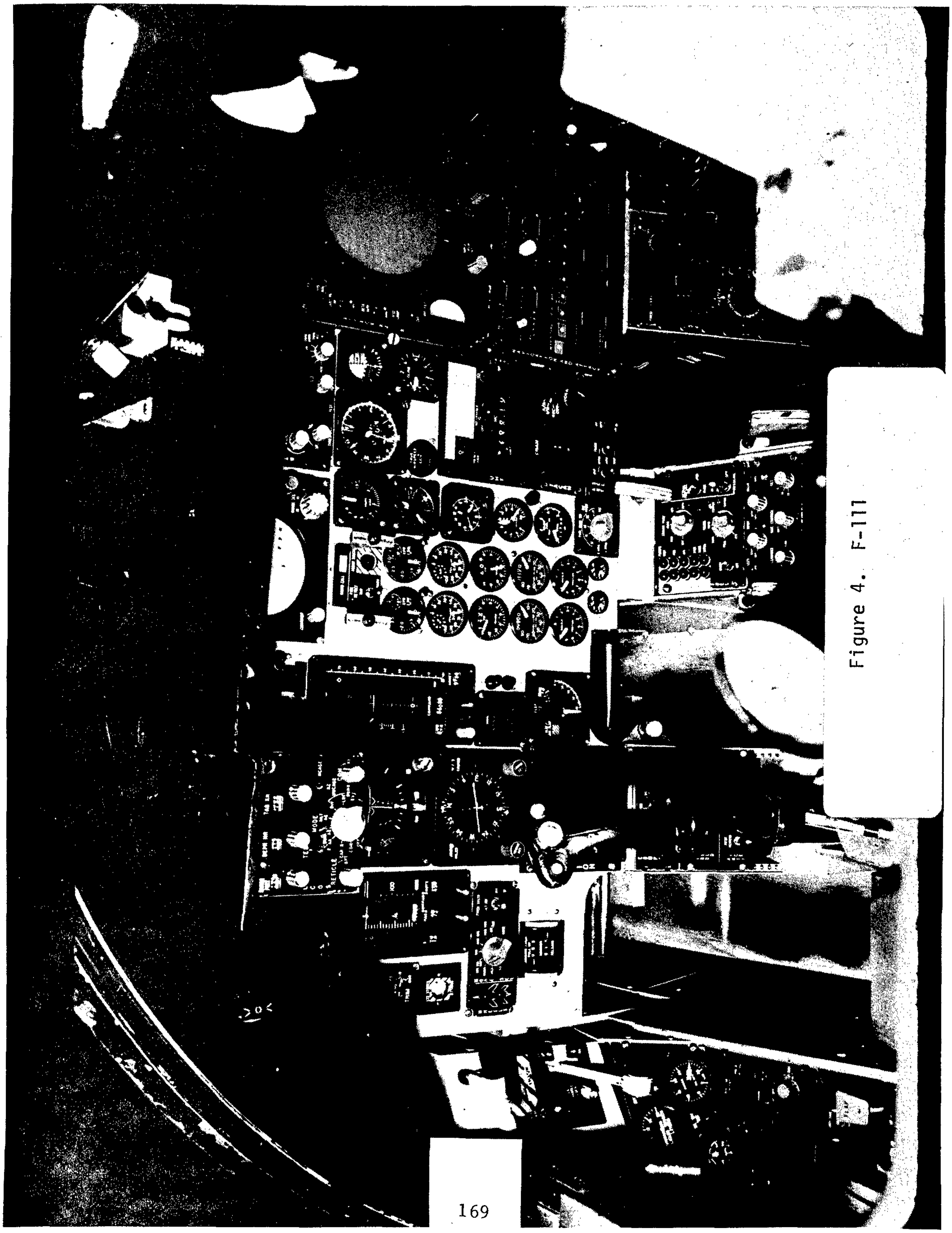
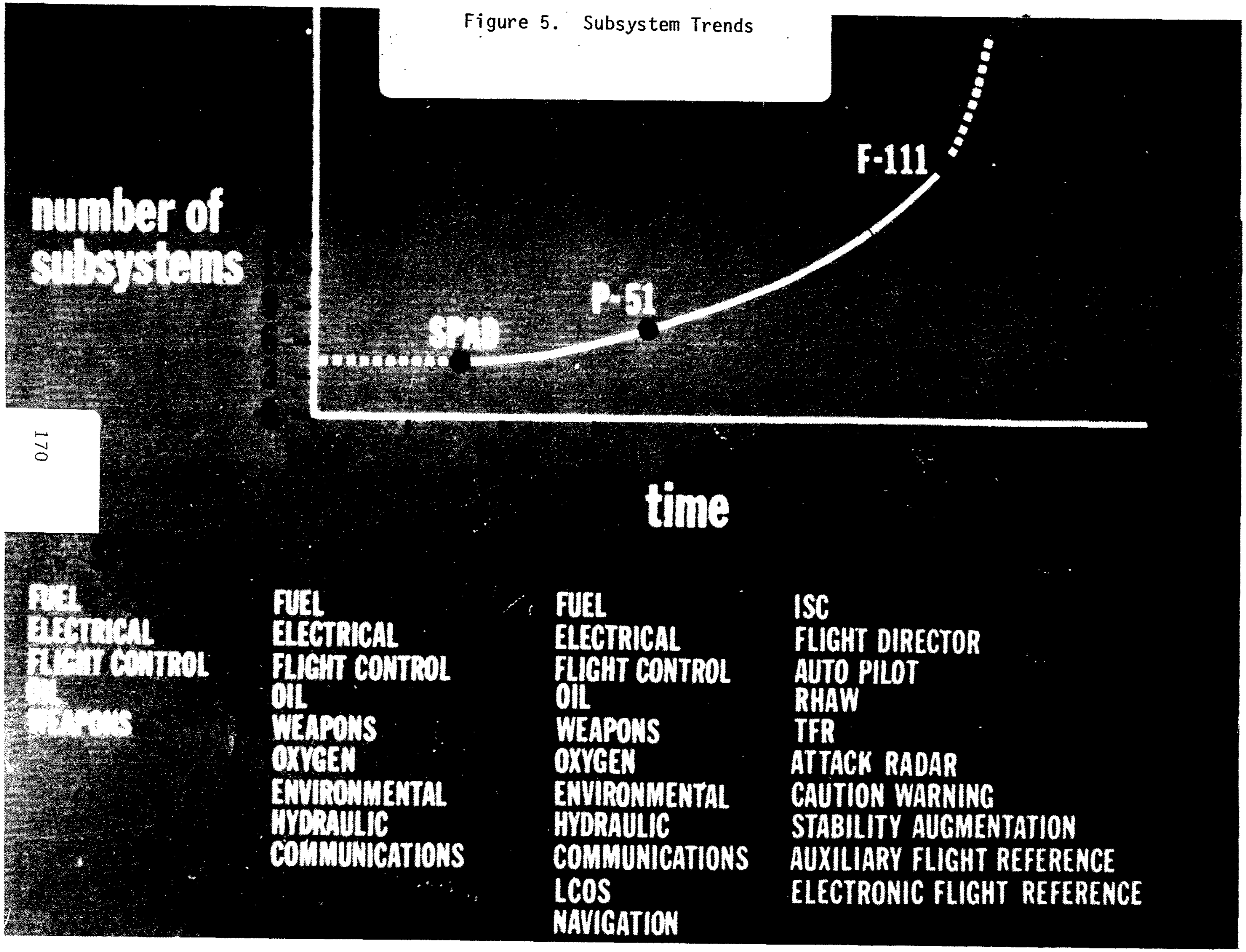


Figure 4. F-111

Figure 5. Subsystem Trends

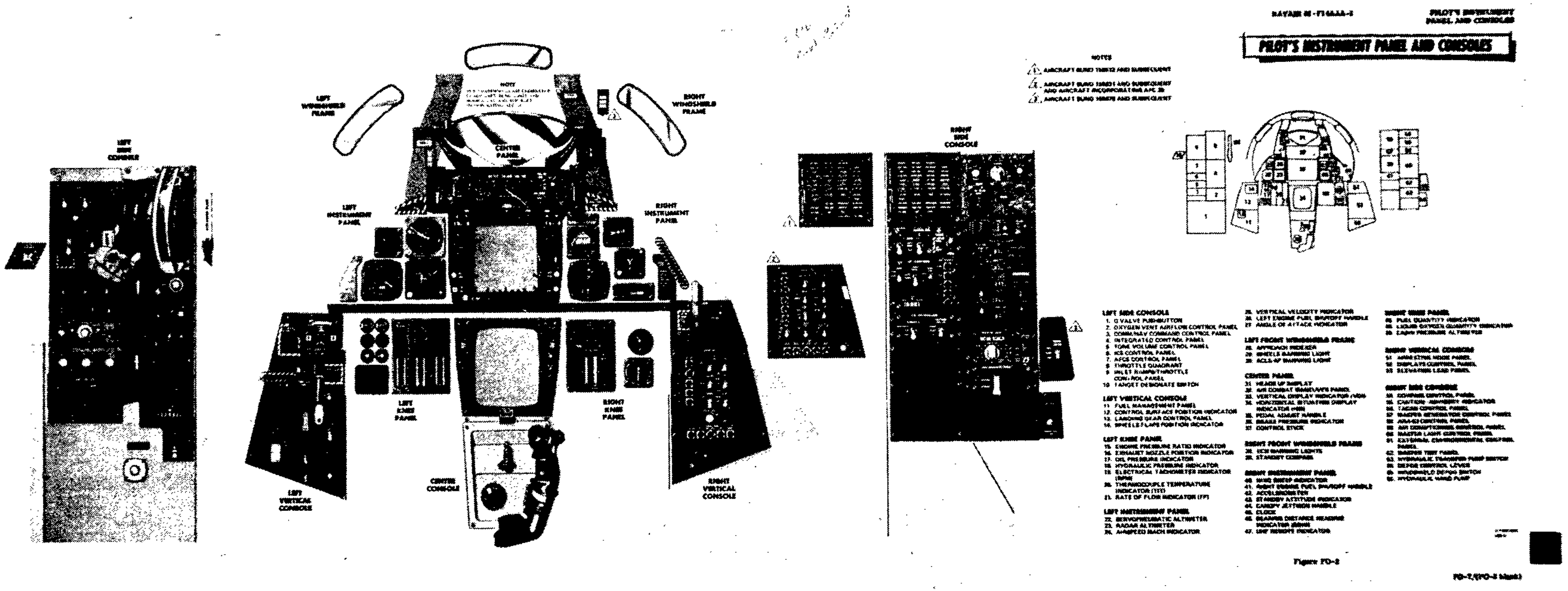


FUEL
ELECTRICAL
FLIGHT CONTROL
OIL
WEAPONS

FUEL
ELECTRICAL
FLIGHT CONTROL
OIL
WEAPONS
OXYGEN
ENVIRONMENTAL
HYDRAULIC
COMMUNICATIONS

FUEL
ELECTRICAL
FLIGHT CONTROL
OIL
WEAPONS
OXYGEN
ENVIRONMENTAL
HYDRAULIC
COMMUNICATIONS
LCOS
NAVIGATION

ISC
FLIGHT DIRECTOR
AUTO PILOT
RHAW
TFR
ATTACK RADAR
CAUTION WARNING
STABILITY AUGMENTATION
AUXILIARY FLIGHT REFERENCE
ELECTRONIC FLIGHT REFERENCE



NAVAIR 01-F14AAA-1

PILOT'S INSTRUMENT PANEL AND CONSOLES

PILOT'S INSTRUMENT PANEL AND CONSOLES

NOTES

1. AIRCRAFT BUNO [illegible] AND SUBSEQUENT
2. AIRCRAFT BUNO [illegible] AND SUBSEQUENT AND AIRCRAFT INCORPORATING AFC [illegible]
3. AIRCRAFT BUNO [illegible] AND SUBSEQUENT

LEFT SIDE CONSOLE
1. G VALVE PUSHBUTTON
2. OXYGEN VENT AIRFLOW CONTROL PANEL
3. COMM/NAV COMMAND CONTROL PANEL
4. INTEGRATED CONTROL PANEL
5. TONE VOLUME CONTROL PANEL
6. ICS CONTROL PANEL
7. AFCS CONTROL PANEL
8. THROTTLE QUADRANT
9. INLET RAMPS/THROTTLE CONTROL PANEL
10. TARGET DESIGNATE SWITCH

LEFT VERTICAL CONSOLE
11. FUEL MANAGEMENT PANEL
12. CONTROL SURFACE POSITION INDICATOR
13. LANDING GEAR CONTROL PANEL
14. WHEELS-FLAPS POSITION INDICATOR

LEFT KNEE PANEL
15. ENGINE PRESSURE RATIO INDICATOR
16. EXHAUST NOZZLE POSITION INDICATOR
17. OIL PRESSURE INDICATOR
18. HYDRAULIC PRESSURE INDICATOR
19. ELECTRICAL TACHOMETER INDICATOR (RPM)
20. THERMOCOUPLE TEMPERATURE INDICATOR (TIT)
21. RATE OF FLOW INDICATOR (FF)

LEFT INSTRUMENT PANEL
22. SERVOPNEUMATIC ALTIMETER
23. RADAR ALTIMETER
24. AIRSPEED MACH INDICATOR
25. VERTICAL VELOCITY INDICATOR
26. LEFT ENGINE FUEL SHUTOFF HANDLE
27. ANGLE OF ATTACK INDICATOR

LEFT FRONT WINDSHIELD FRAME
28. APPROACH INDEXER
29. WHEELS WARNING LIGHT
30. ACLS/AP WARNING LIGHT

CENTER PANEL
31. HEAD UP DISPLAY
32. AIR COMBAT MANEUVER PANEL
33. VERTICAL DISPLAY INDICATOR (VDI)
34. HORIZONTAL SITUATION DISPLAY INDICATOR (HSD)
35. PEDAL ADJUST HANDLE
36. BRAKE PRESSURE INDICATOR
37. CONTROL STICK

RIGHT FRONT WINDSHIELD FRAME
38. ECM WARNING LIGHTS
39. STANDBY COMPASS

RIGHT INSTRUMENT PANEL
40. WING SWEEP INDICATOR
41. RIGHT ENGINE FUEL SHUTOFF HANDLE
42. ACCELEROMETER
43. STANDBY ATTITUDE INDICATOR
44. CANOPY JETTISON HANDLE
45. CLOCK
46. BEARING DISTANCE HEADING INDICATOR (BDHI)
47. UHF REMOTE INDICATOR

RIGHT KNEE PANEL
48. FUEL QUANTITY INDICATOR
49. LIQUID OXYGEN QUANTITY INDICATOR
50. CABIN PRESSURE ALTIMETER

RIGHT VERTICAL CONSOLE
51. ARRESTING HOOK PANEL
52. DISPLAYS CONTROL PANEL
53. ELEVATION LEAD PANEL

RIGHT SIDE CONSOLE
54. COMPASS CONTROL PANEL
55. CAUTION ADVISORY INDICATOR
56. TACAN CONTROL PANEL
57. MASTER GENERATOR CONTROL PANEL
58. ARA-63 CONTROL PANEL
59. AIR CONDITIONING CONTROL PANEL
60. MASTER LIGHT CONTROL PANEL
61. EXTERNAL ENVIRONMENTAL CONTROL PANEL
62. MASTER TEST PANEL
63. HYDRAULIC TRANSFER PUMP SWITCH
64. DEFOG CONTROL LEVER
65. WINDSHIELD DEFOG SWITCH
66. HYDRAULIC HAND PUMP

Figure FO-2

FO-2/(FO-3 blank)

Figure 6. F-14

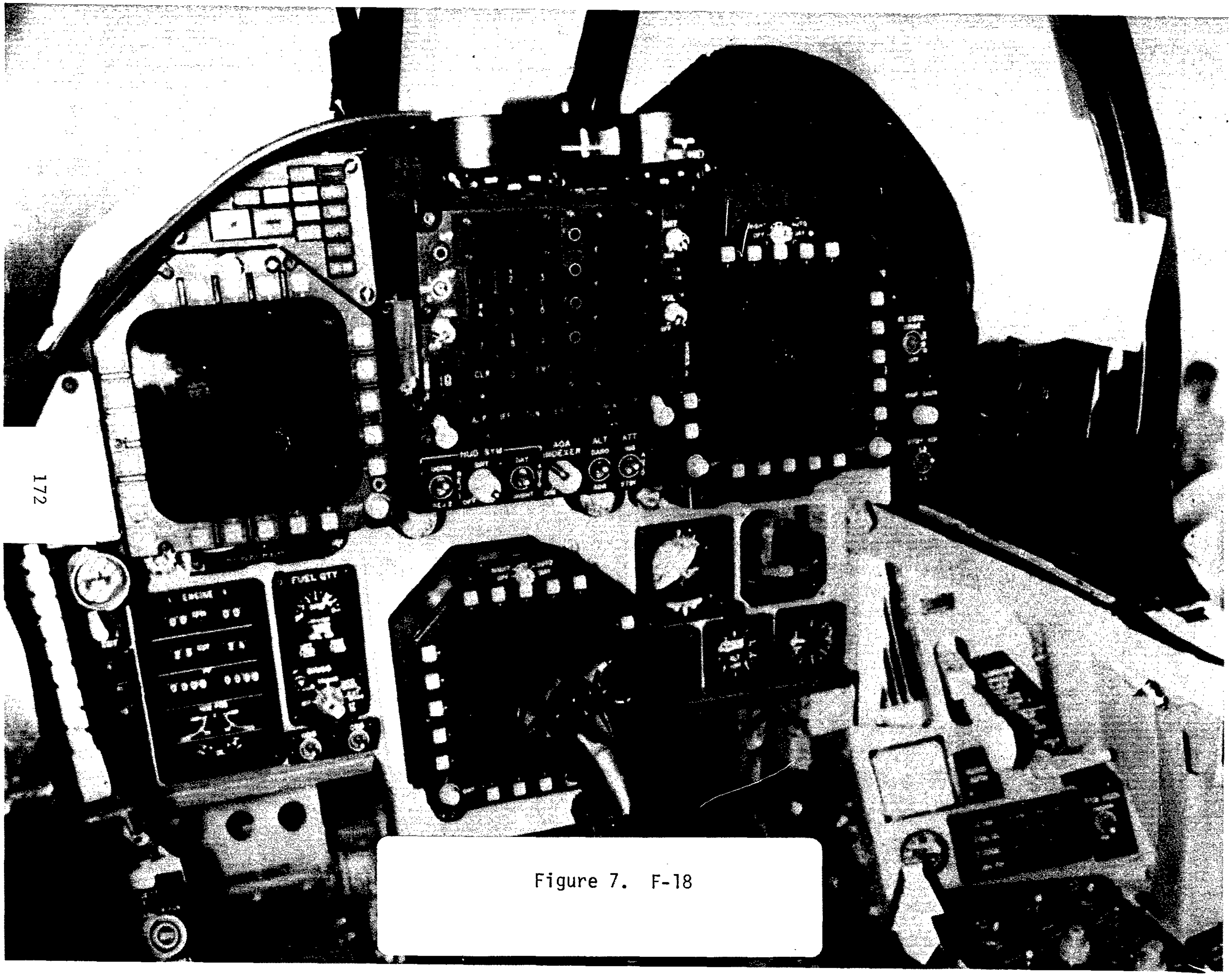


Figure 7. F-18

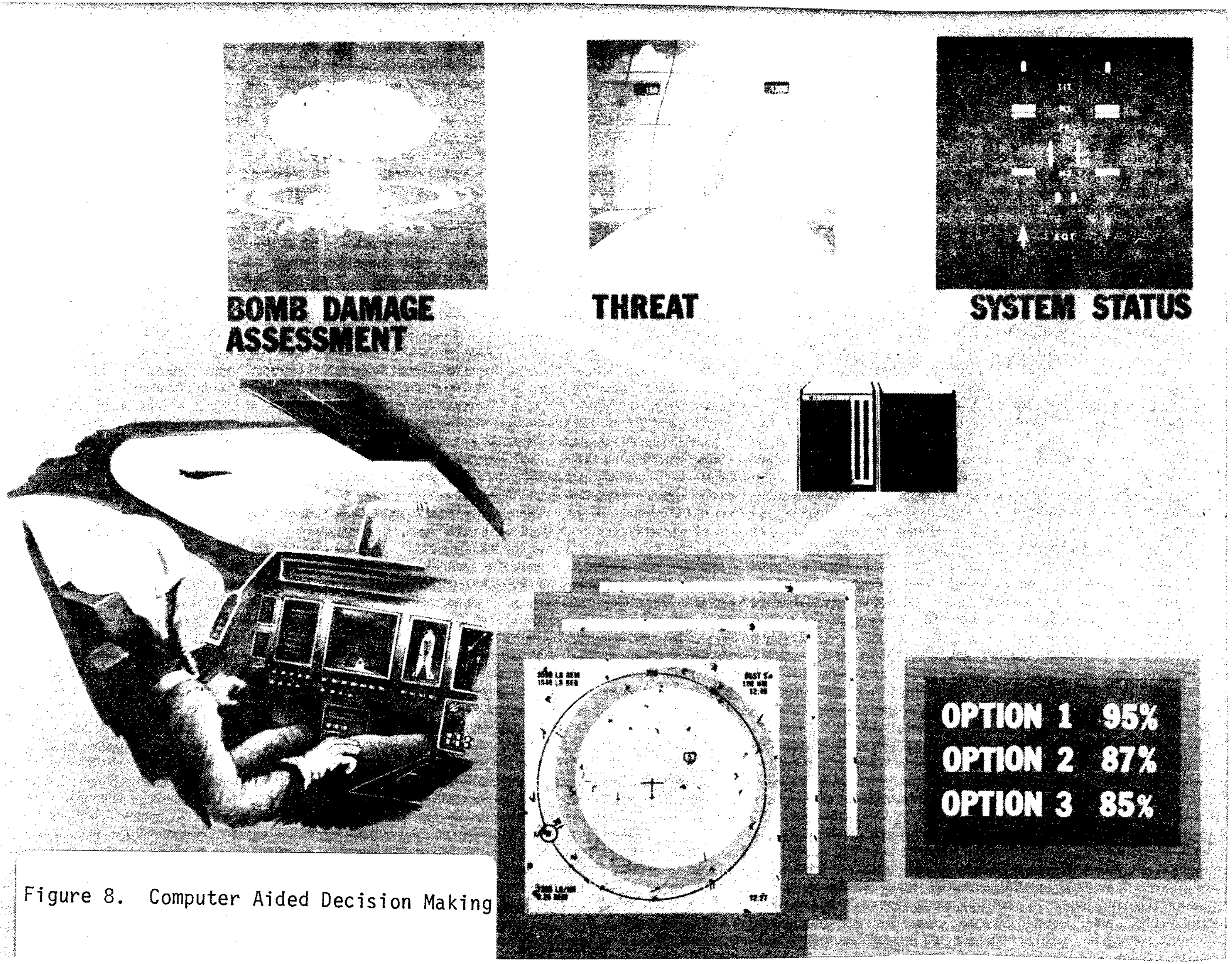


Figure 8. Computer Aided Decision Making

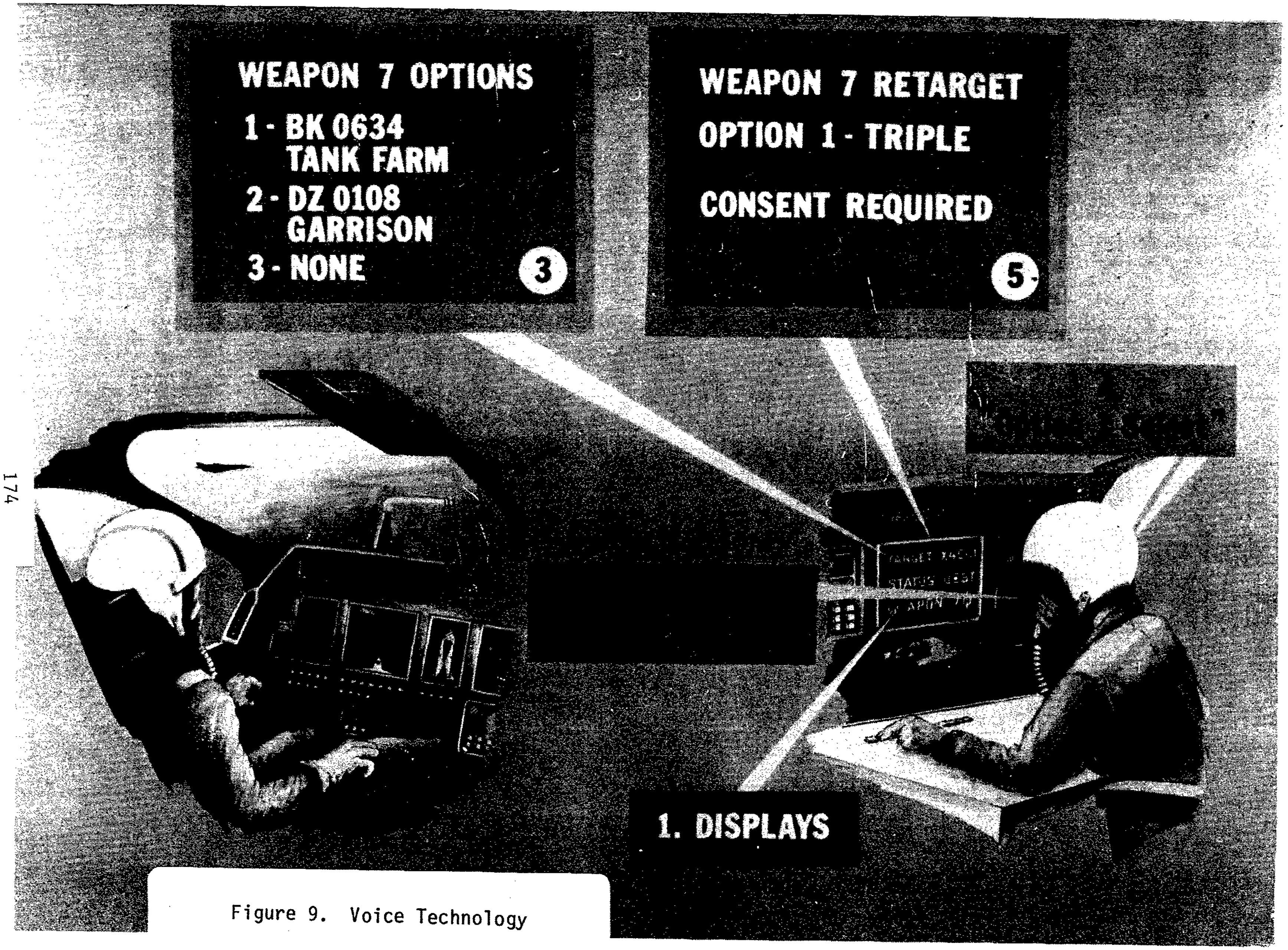


Figure 9. Voice Technology

RECOGNITION TASK CANDIDATES

- CHECKLIST RECALL
- PERFORMANCE CHART RECALL, MAP DATA
- RADIO FREQUENCY CHANGE (DURING TO/LAND)
- ALTIMETER SETTINGS (PRESSURE, RADAR)
- IFF CODES
- CLEARANCE PLANE SETTING (TERRAIN AVIODANCE)
- TERRAIN DISPLAY MODE SELECTION
- RECORD KEEPING TASKS
- HSI INTERACTIONS (COMMAND HEADING)

81V63

Figure 10. Recognition Task Candidates

GENERATION TASK CANDIDATES

- ALTITUDE CALLS (LANDING)
- MASTER CAUTION PANEL MESSAGES (WARNINGS)
- AIRSPEED CALLS (TO/LANDING)
- FLAP POSITION
- SLANT RANGE CALLS (IN LOW-LEVEL ENTRY)
- "TO GO" CALLS IN SECONDS (BOMBING RUNS)

Figure 11. Generation Task Candidates

81V63

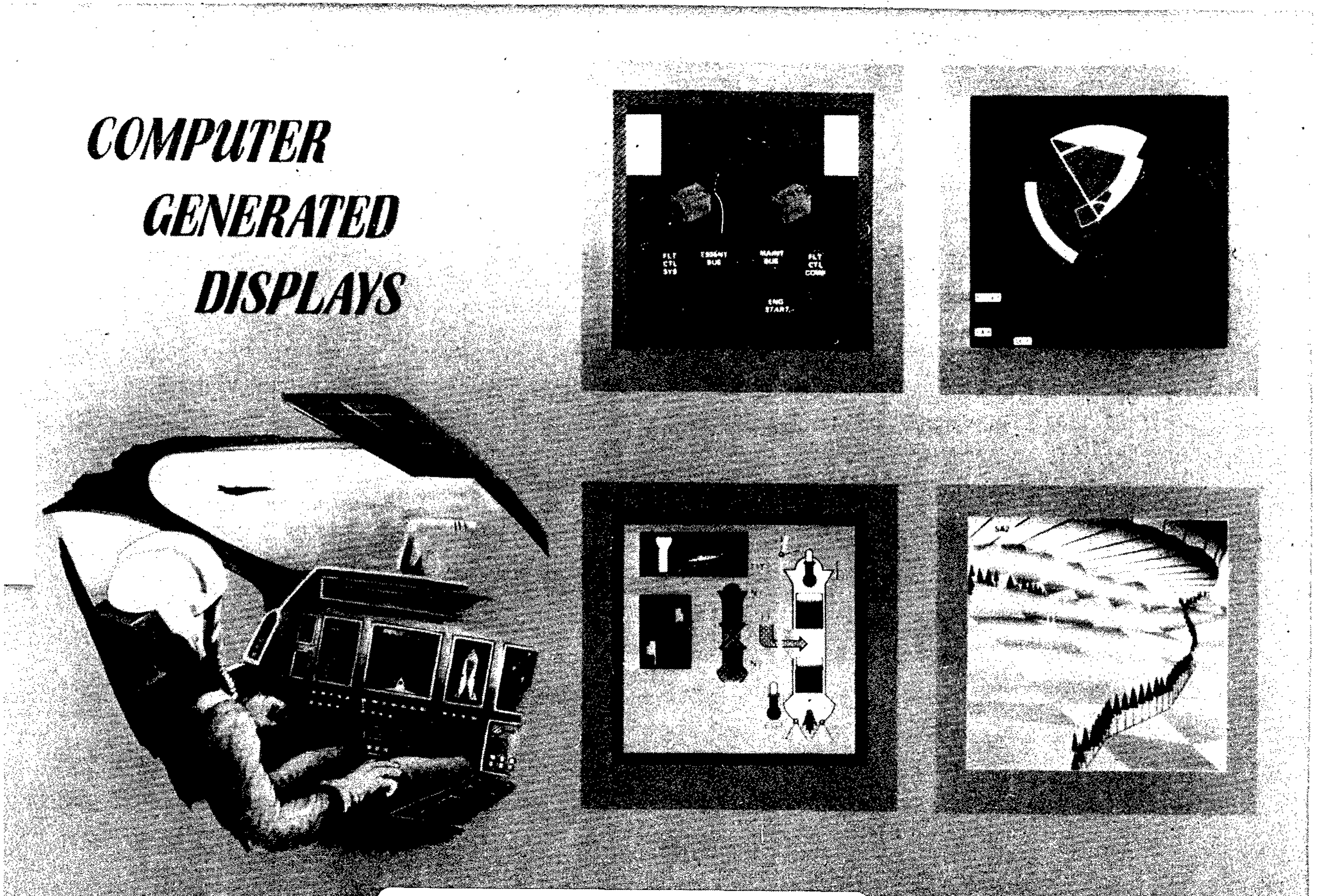


Figure 12. Computer Generated Displays

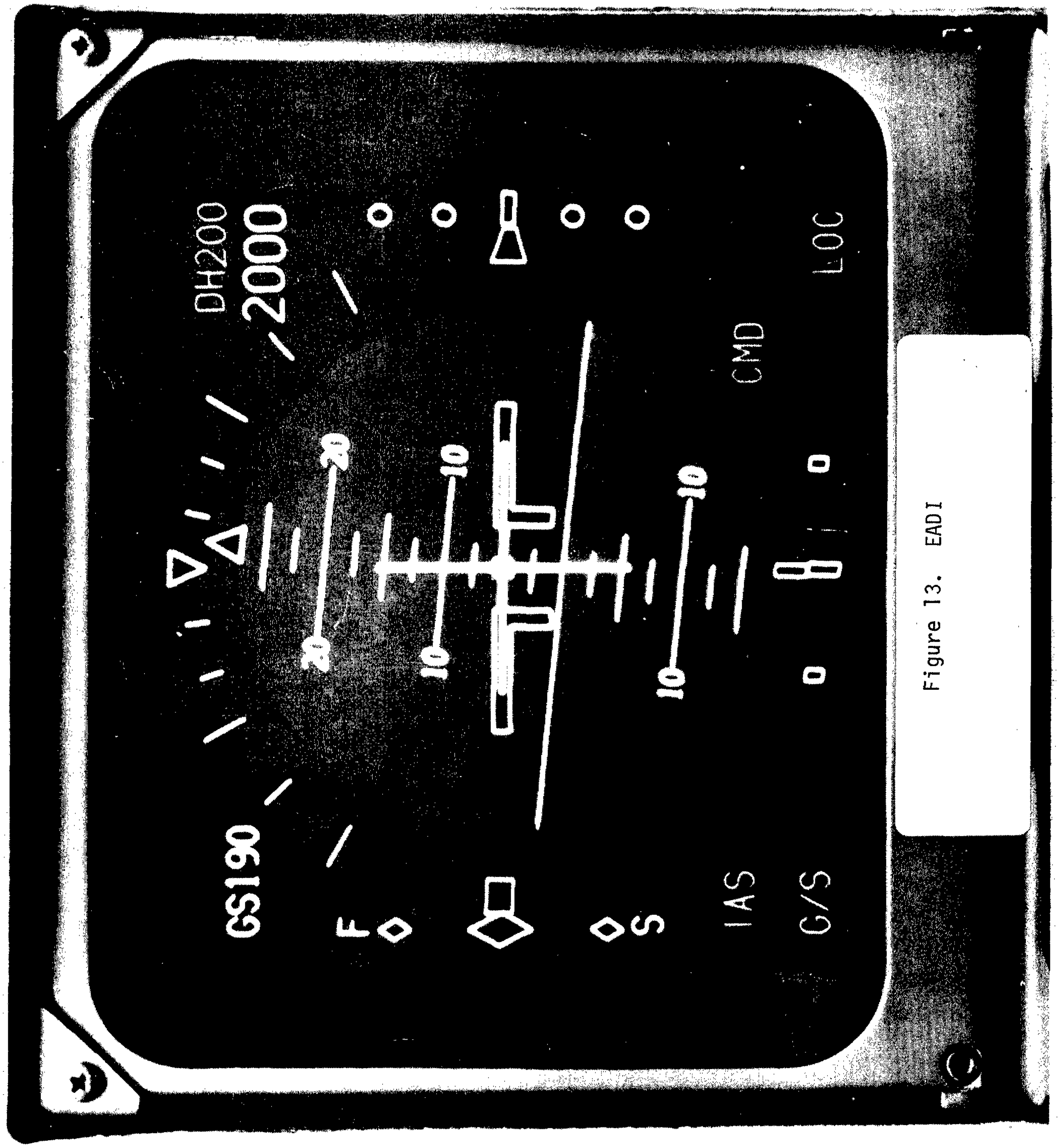


Figure 13. EADI

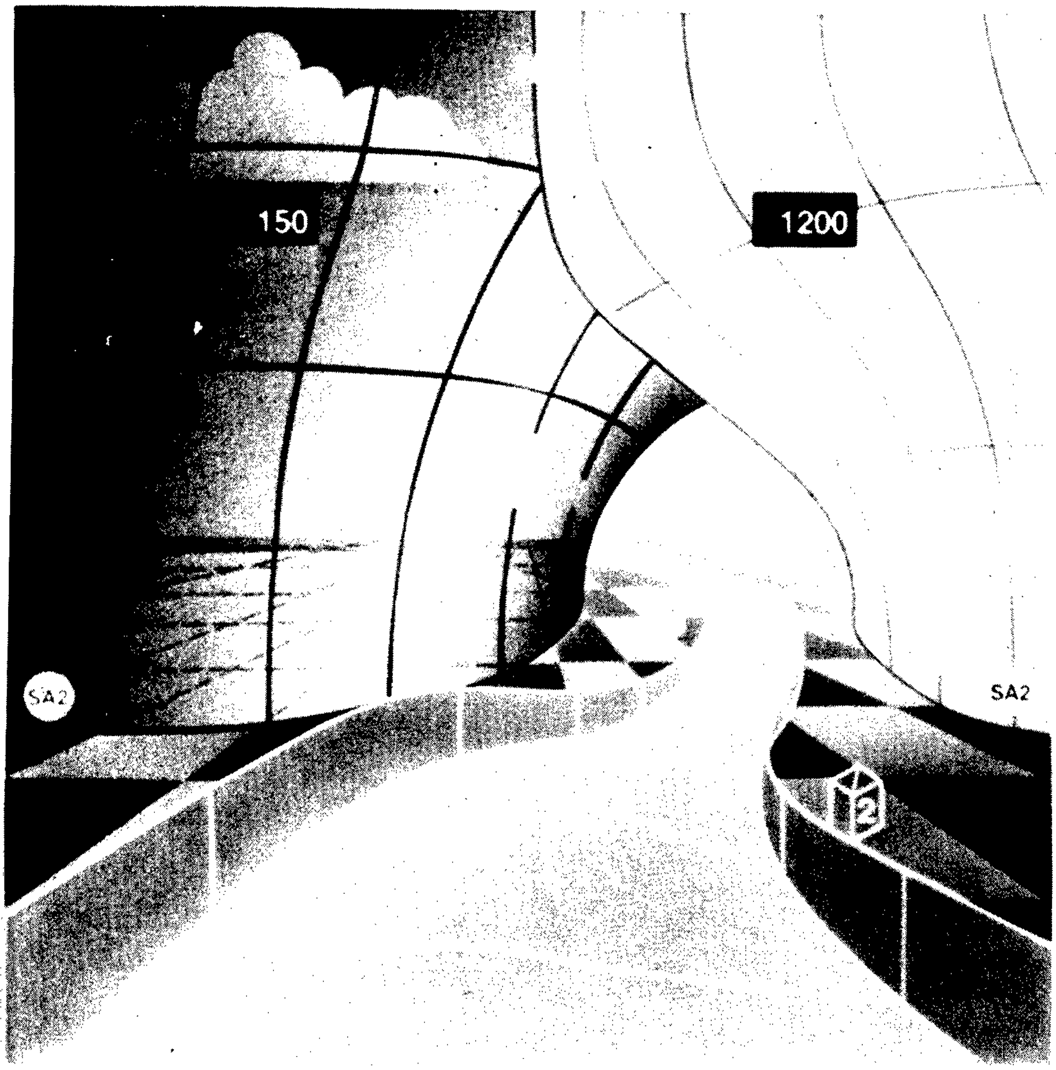


Figure 14. Primary Flight and
Tactical Situation Display

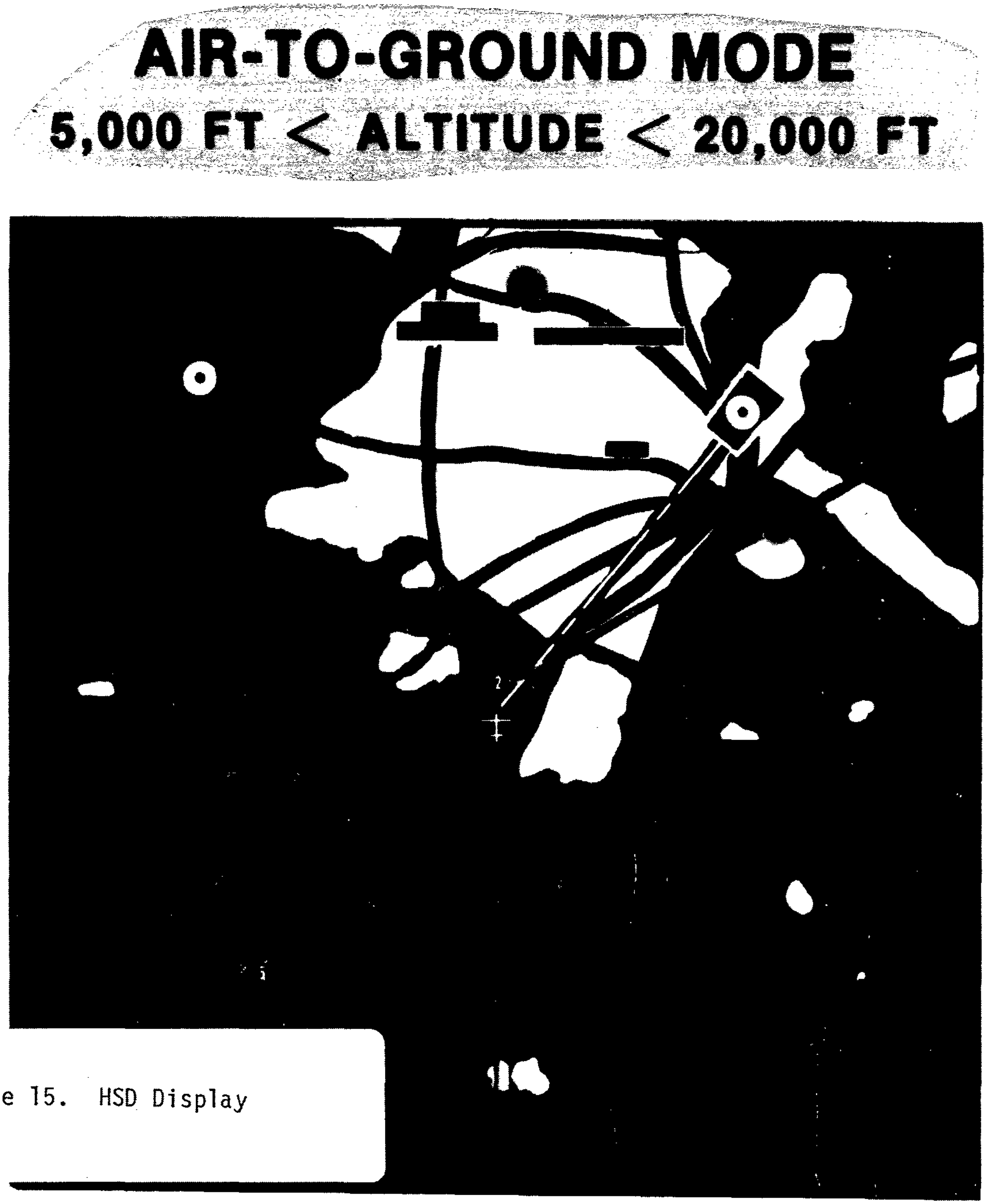
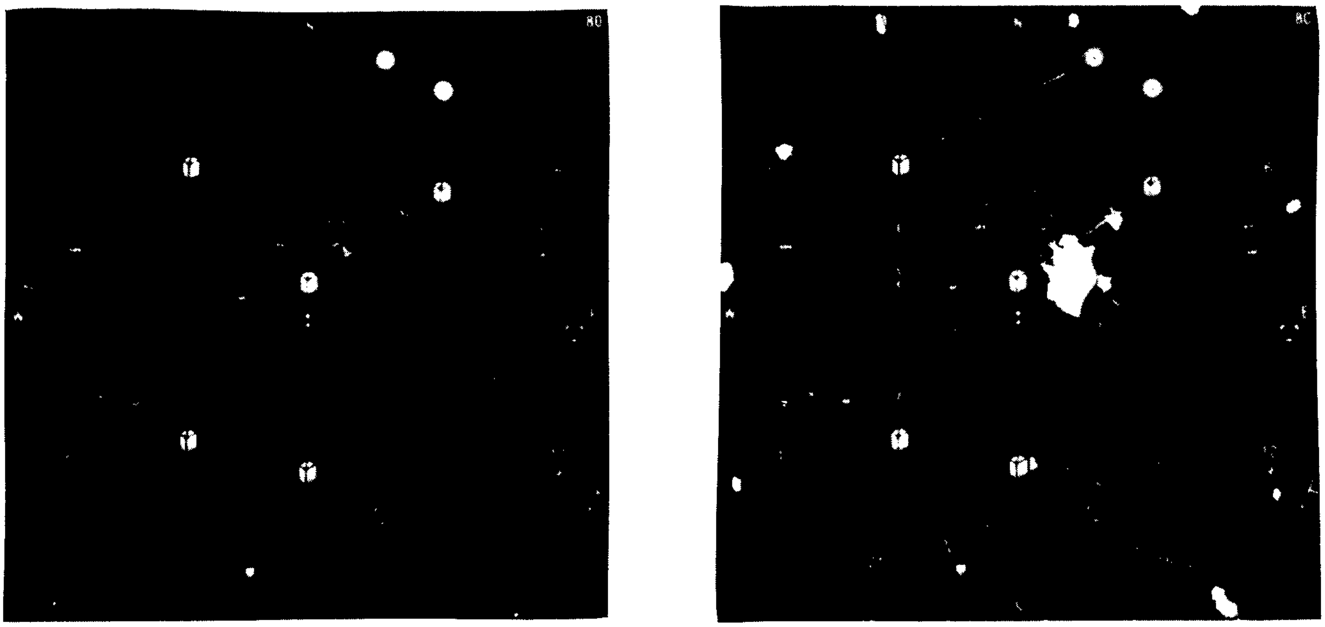


Figure 15. HSD Display

NAVIGATION MODE

ALTITUDE > 20,000 FT



DEFAULT DATA

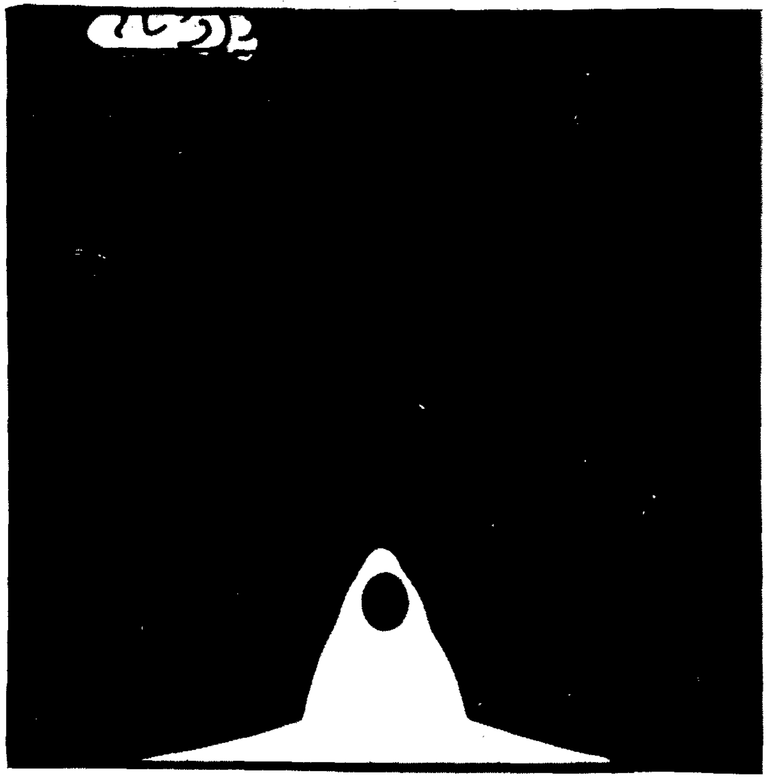
EXAMPLE DATA SUPPLEMENTS

Figure 16. HSD Display

Figure 17. Transition Display

PRIMARY FLIGHT · TACTICAL SITUATION

VIEW POINT
MOVED OUT
SIDE OF
COCKPIT



SCALE ≈ 1:1

INITIAL SHIFT

HEAD UP - HEAD DOWN
1:1 SCALE - ZOOM TO PROGRESSIVELY
SMALLER SCALES

GP01-0870-90a

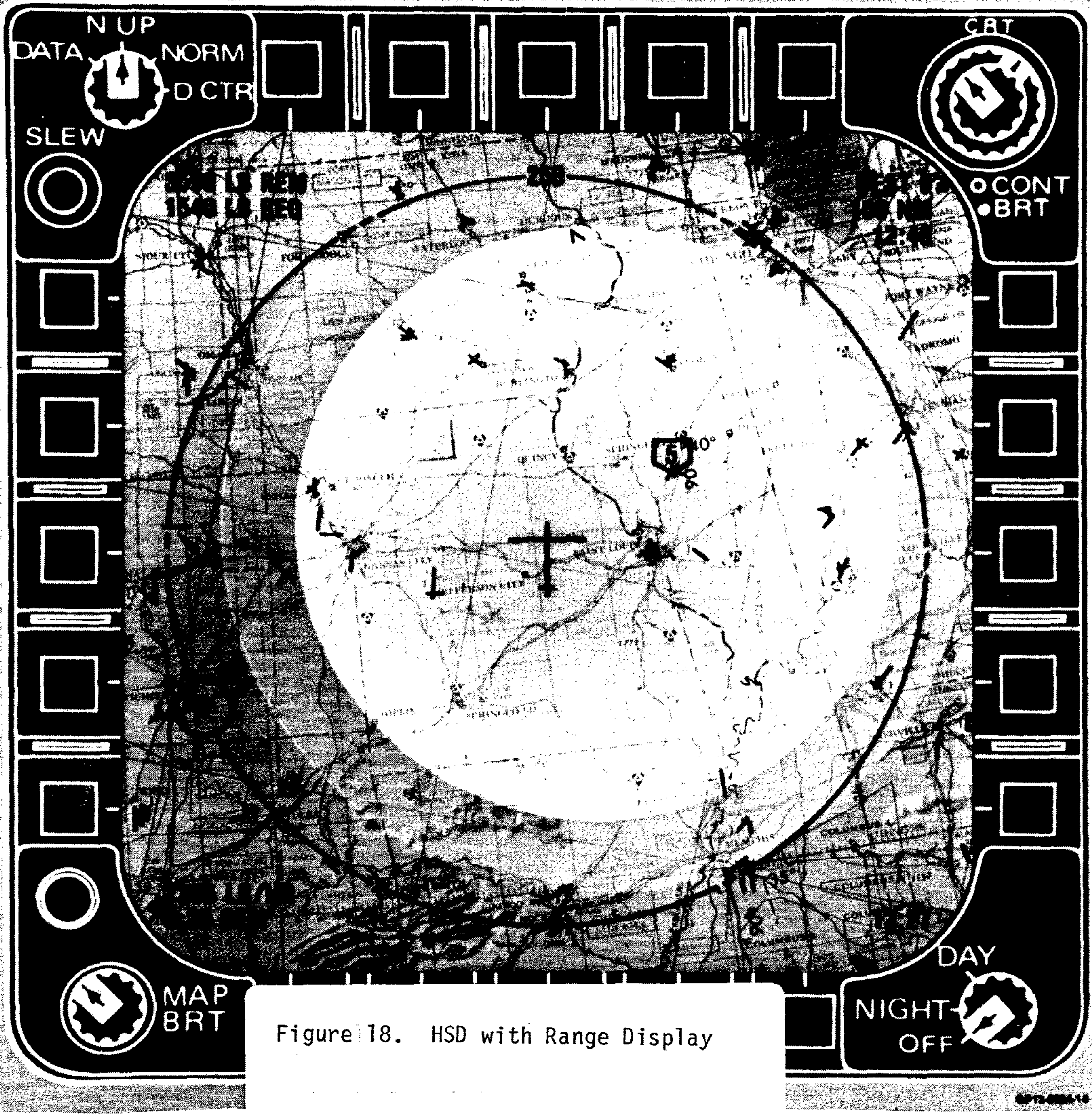


Figure 18. HSD with Range Display

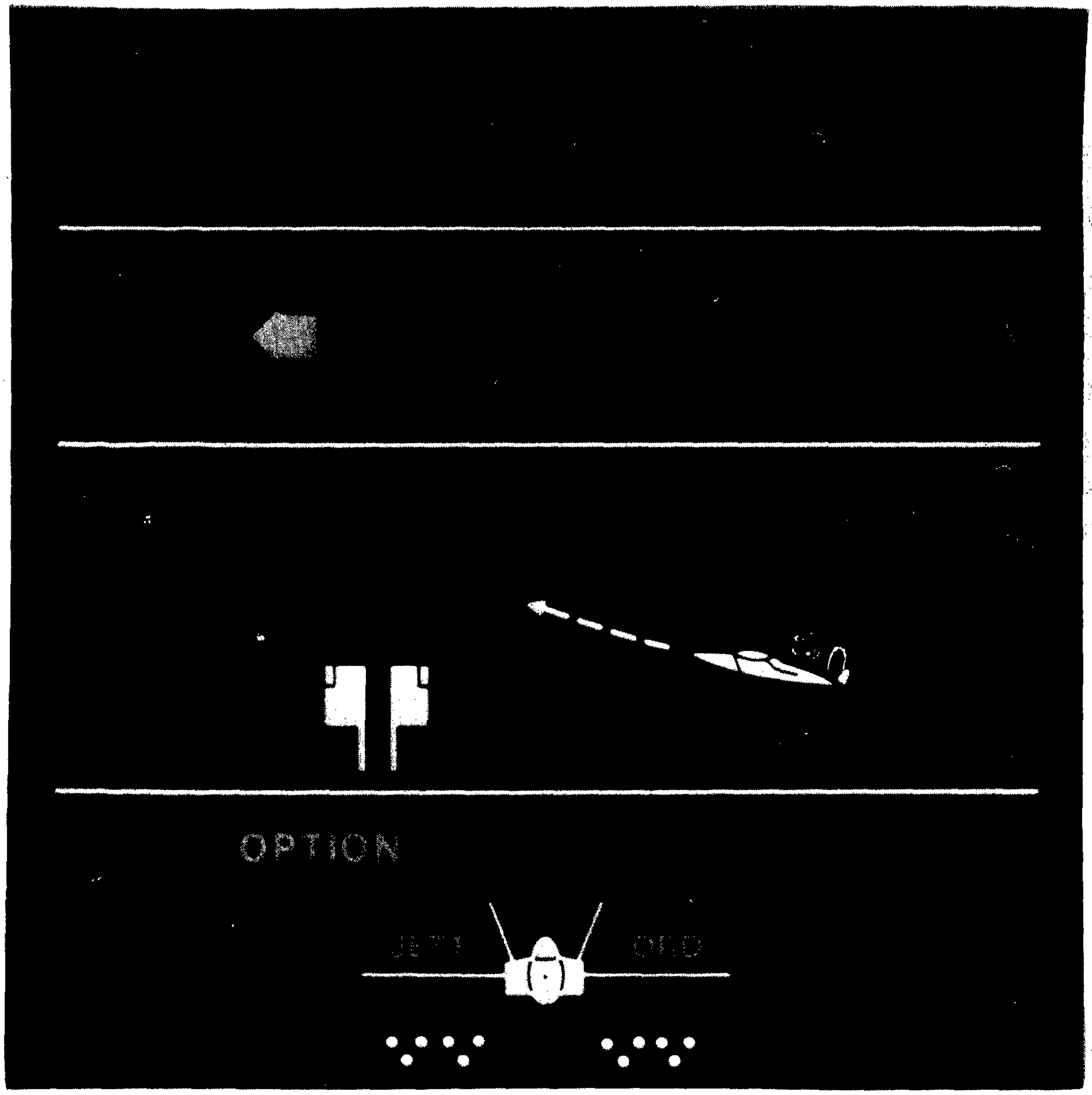


Figure 19. Fuel Endurance Procedures on MPD Display

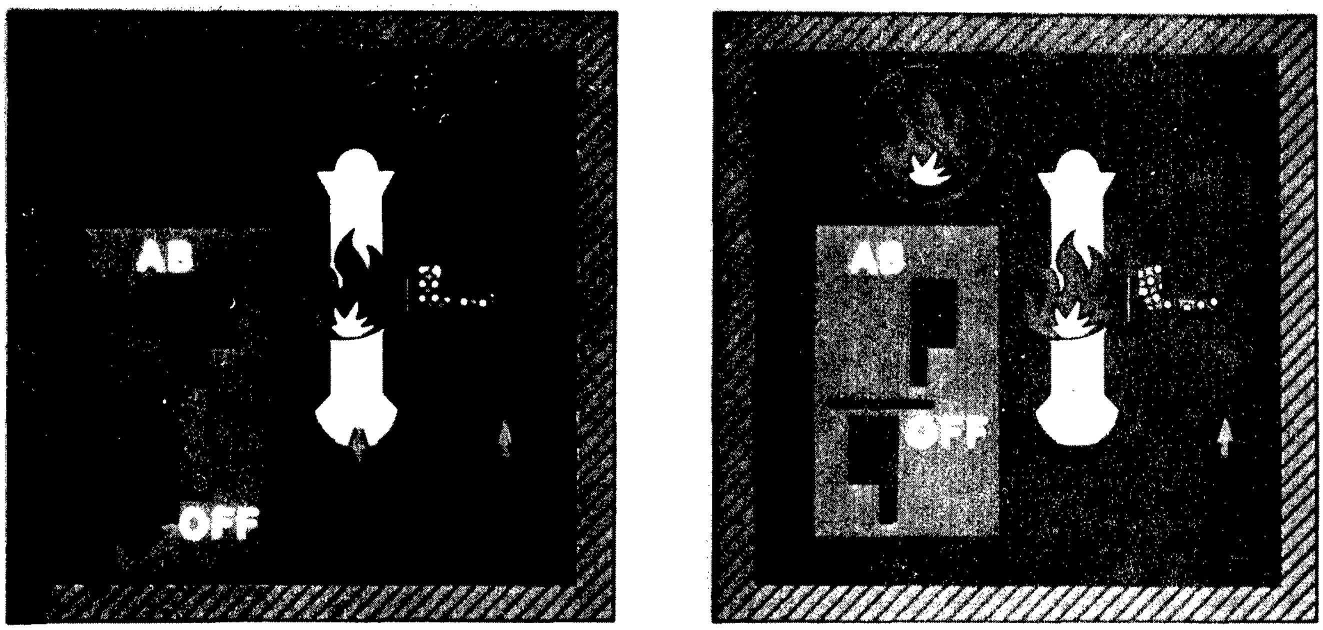
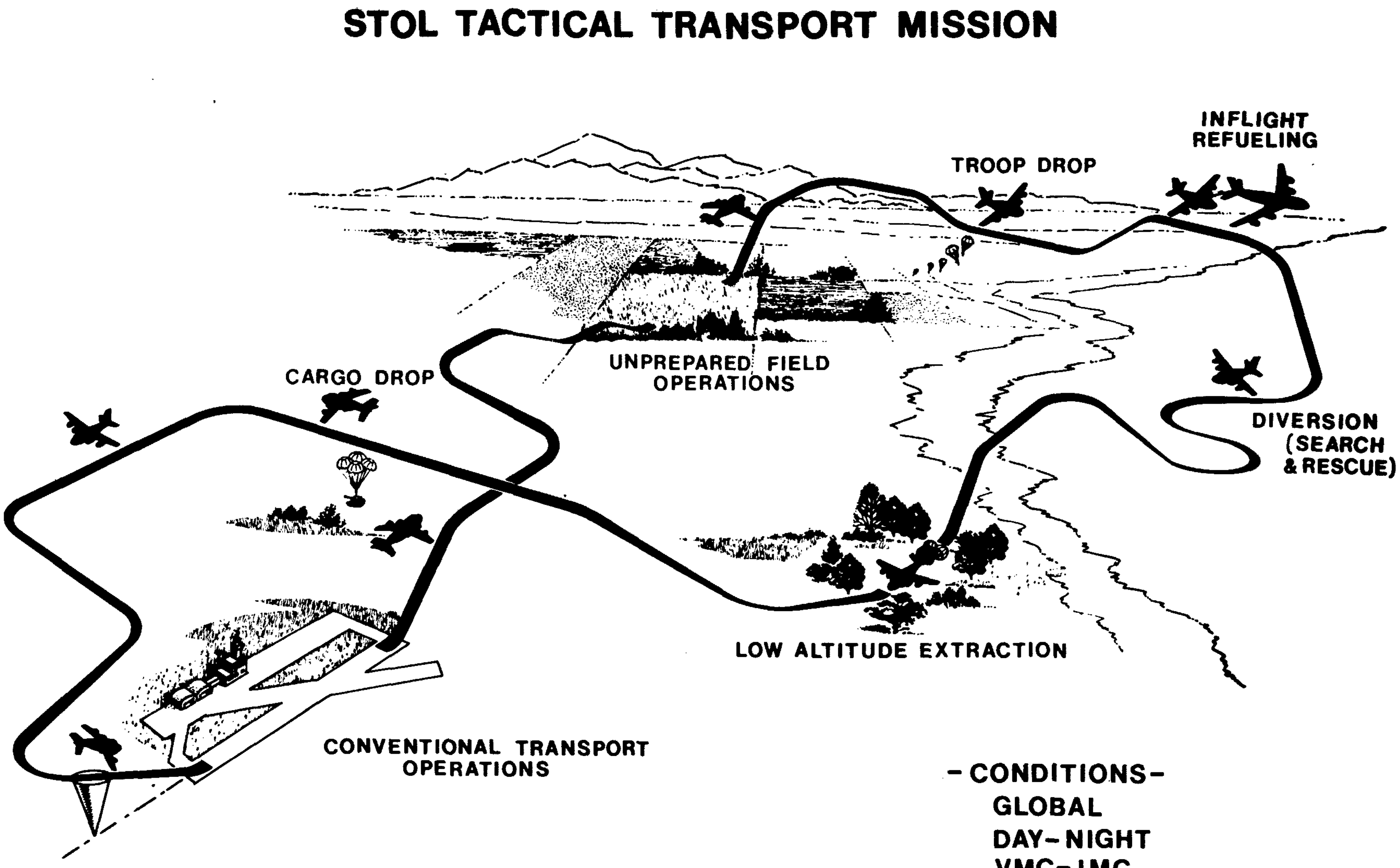


Figure 20. Engine Fire Emergency Procedures on MPD Display



- CONDITIONS -
GLOBAL
DAY-NIGHT
VMC-IMC
THREAT ENVIRONMENT

Figure 21. Mission Scenario

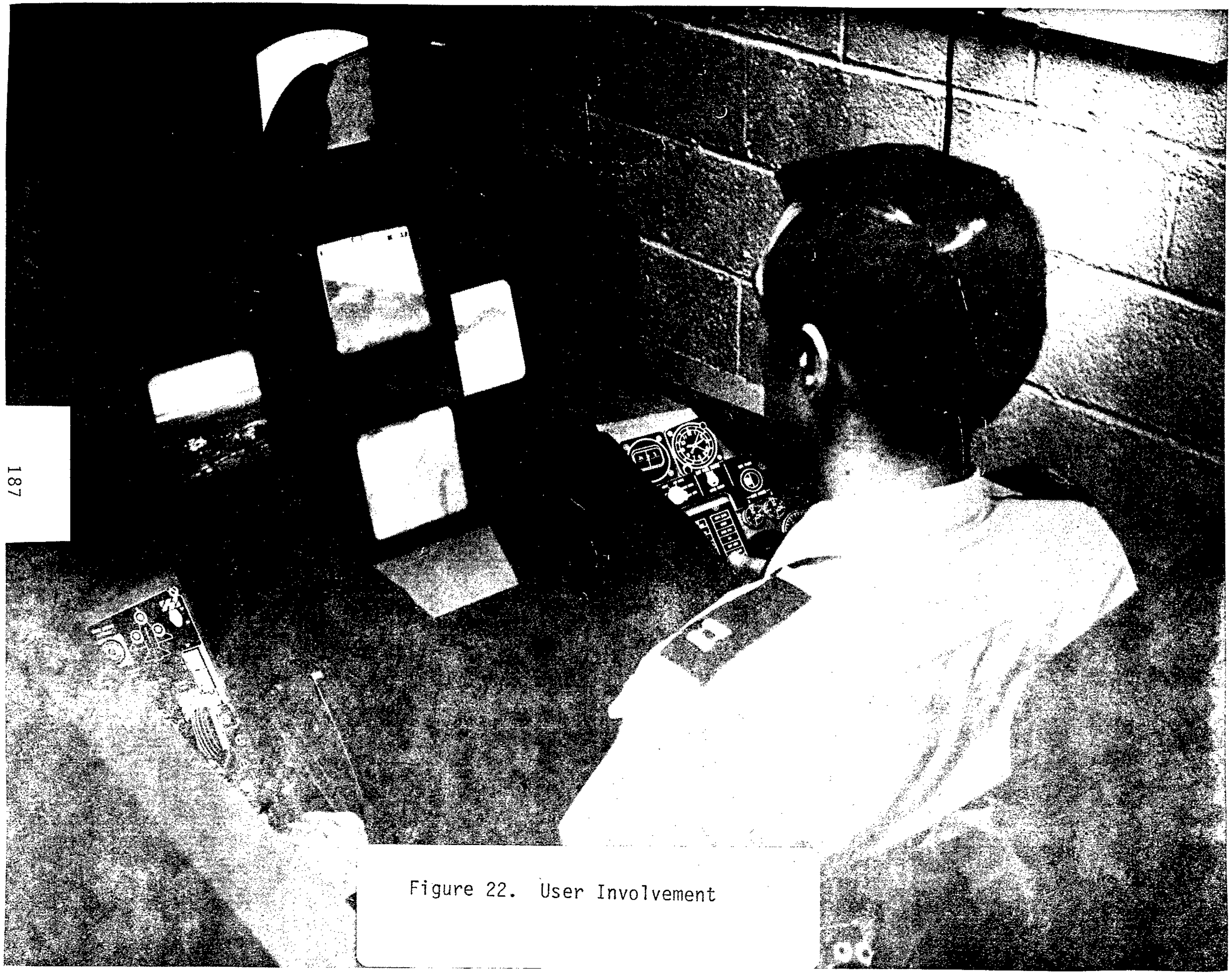
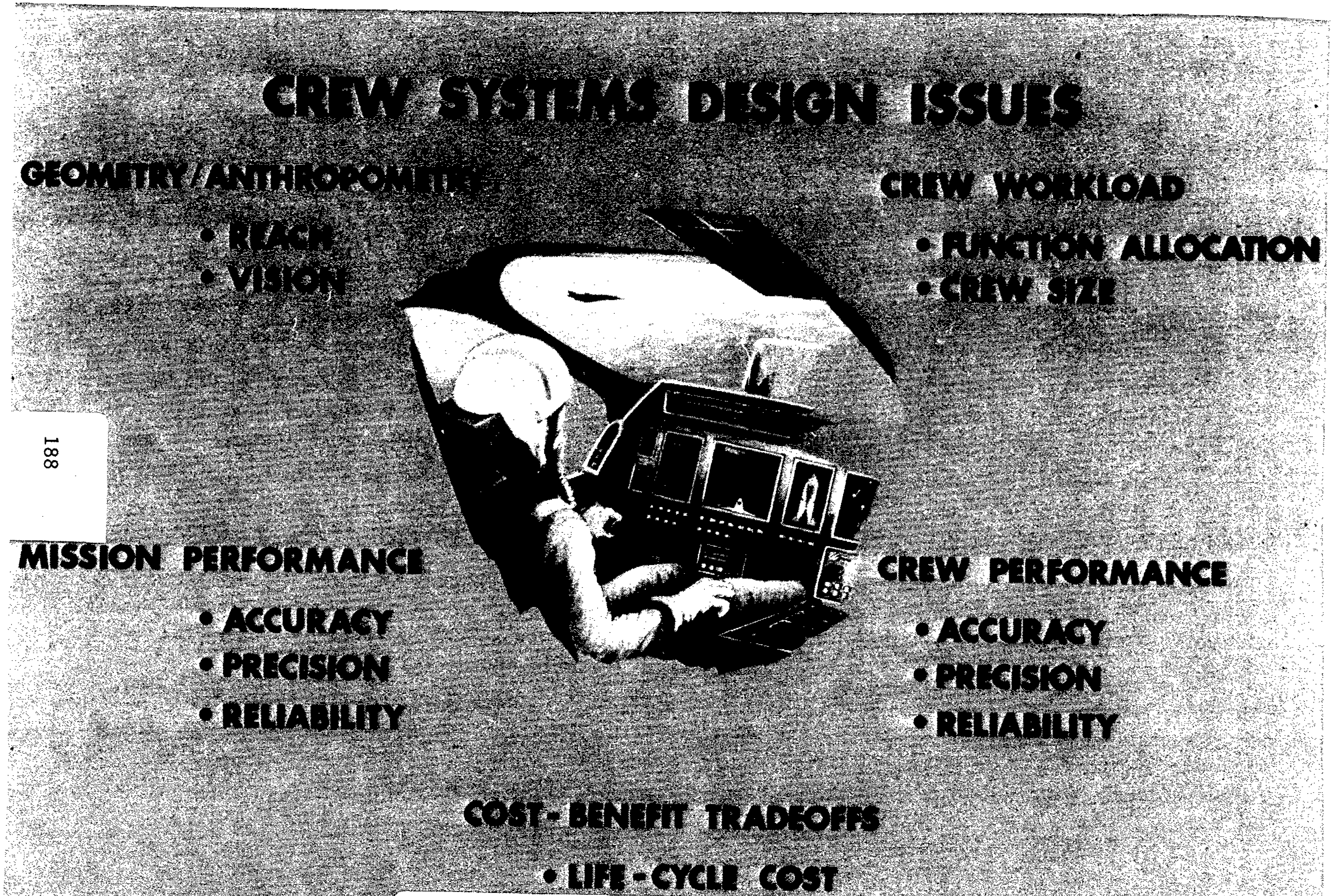


Figure 22. User Involvement



CREW SYSTEMS DESIGN ISSUES

GEOMETRY/ANTHROPOMETRY
- REACH
- VISION

CREW WORKLOAD
- FUNCTION ALLOCATION
- CREW SIZE

MISSION PERFORMANCE
- ACCURACY
- PRECISION
- RELIABILITY

CREW PERFORMANCE
- ACCURACY
- PRECISION
- RELIABILITY

COST-BENEFIT TRADEOFFS
- LIFE-CYCLE COST

Figure 23. Crew Systems Design Issue

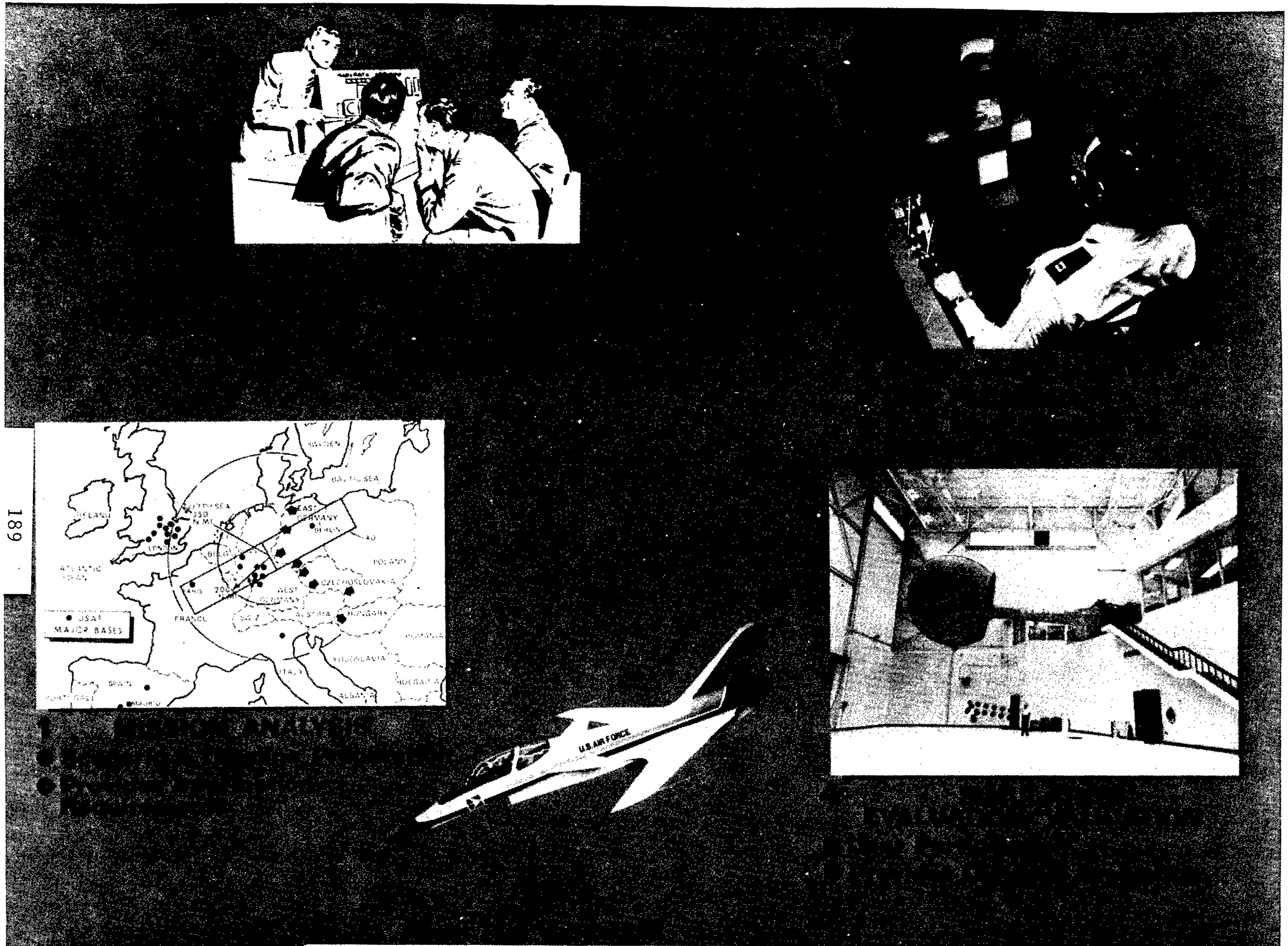


Figure 24. Crew Systems Design Process

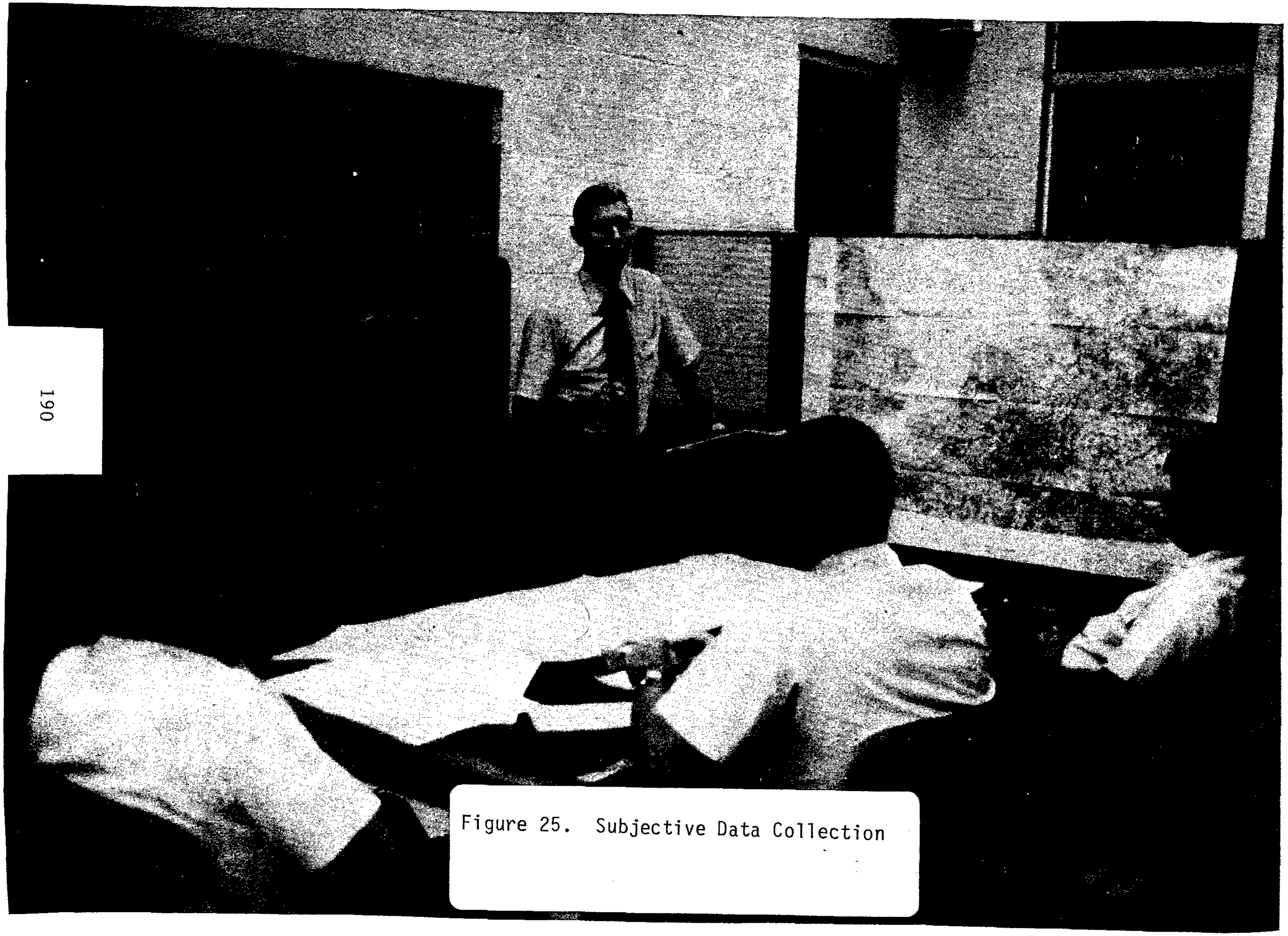
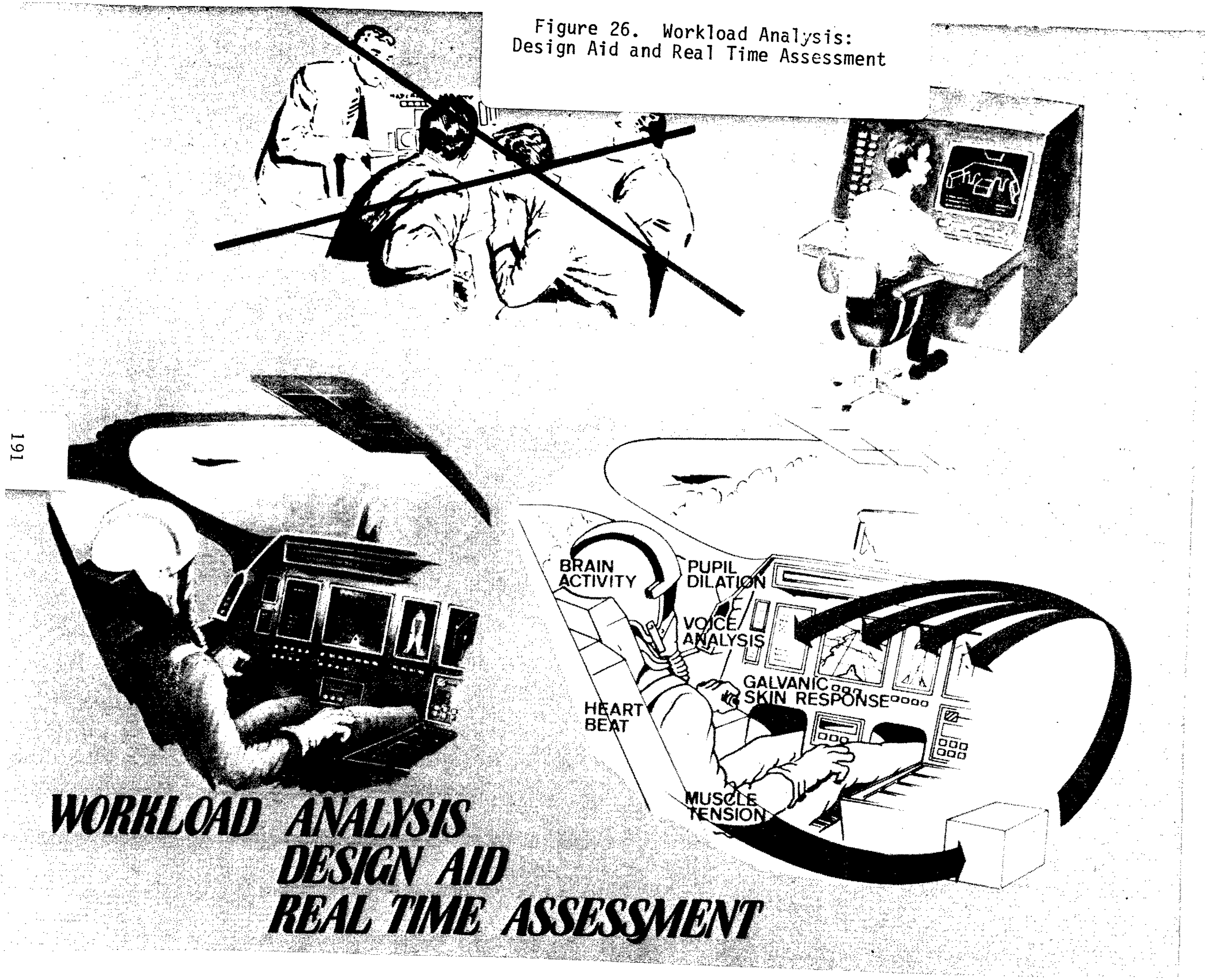


Figure 25. Subjective Data Collection

Figure 26. Workload Analysis:
Design Aid and Real Time Assessment



RELATIONSHIP OF THE FLYING QUALITIES SPECIFICATION TO TASK PERFORMANCE

Frank L. George
David J. Moorhouse

Flying Qualities Group, Flight Dynamics Laboratory
Air Force Wright Aeronautical Laboratories

SUMMARY

The original military specification was a definitive set of flight test maneuver instructions intended to demonstrate acceptable stability and control (and, therefore, handling qualities) characteristics, By 1969, the specification had evolved to better fit the design process by giving acceptable ranges of stability and control parameters, and reliability goals. It appears to some observers that, by this same time, the design process had been segmented so there was a polarization between the aerodynamics and performance people on one side and the control designers on the other. As a result, there seems to be a competition - the aerodynamicists challenging the control people with increasingly exotic configurations and larger performance envelopes; the flight control designers countering with more complex control laws and feedback structures. This fragmented approach leads the designers away from the essence of our flying qualities definition - the eventual integration of pilot and airplane characteristics for some useful task. Thus it seems capability is sometimes put into an airplane just because it's possible rather than because it's useful.

Of course, the need to properly interface the pilot and airplane was recognized long ago. Development of the understanding and tools to accomplish the feat has been, and continues to be, slow and difficult. Notable developments have been the servo and optimal pilot models, the Cooper-Harper and related subjective evaluation scales and extensive computational aids for dynamic analysis and data processing. Numerous design methods and criteria have been proposed and tested using these basic results, only a sampling of which will be discussed in this paper. Interest in pilot identification with regard to flying qualities arises because the use of experimental methods in criteria development as well as in design and evaluation leads to the need for isolating the pilot component of closed loop responses in order to define vehicle requirements or dynamics.

SPECIFICATION BACKGROUND

The subject of flying qualities has been discussed under one title or another for as long as people have tried to make the airplane a useful device. Figure 1 traces the early evolution of military requirements for what we now call flying qualities. The quote from the 1907 contract for a Wright flyer was essentially the entire specification of "flying qualities". The human piloting aspects were recognized, however, through the statement "It should be sufficiently simple in its construction and operation to permit an intelligent man to become proficient in its use within a reasonable length of time." The first formal specification, AAF-C-1815, was essentially a collection of required flight test maneuver instructions to demonstrate stability and control limits. An example of the requirements is shown in Figure 2. As can be seen from the maneuver description, this requirement was as much a demonstration of performance capability as it was control, and the only consideration for the pilot was the limit on stick force. Since Class III airplanes in that specification included fighters, attack planes and dive bombers some relationship can be seen between the pullup requirements and mission needs, but the demonstration maneuver doesn't necessarily represent the way a pilot would use the pullup capability operationally.

Reference 1 traces the development of appreciation for the pilot's influence on the controlled airplane's dynamic properties, and of analytic methods to describe and study the pilot-airplane system. (Ron Anderson's paper at this workshop also gives a detailed discussion of pilot models).

A significant change occurred with the introduction of pilot-vehicle analysis methods and the use of rating scales to quantify the pilot's subjective assessment of flying qualities. What followed was a period of specialization in flying qualities research along two main tracks which, at times, have even appeared to diverge. One track emphasized the interaction between the pilot dynamics and airplane dynamics; pilot models were developed more as a research activity with little practical acceptance. The other track concentrated on defining open loop characteristics which would satisfy closed loop requirements. Augmentation concepts were developed to provide reasonable stability and effective control over larger and larger performance envelopes.

What was the effect on requirements and specifications? A'Harrah and Woodcock (Ref 2) assert the current military specification requirements (Ref 3) "are largely predicted on his [the pilot's] need to perform fine tracking". They further assert the specification has evolved from a demonstration document to one primarily useful as a design guide. Many requirements in MIL-F-8785B were stated in terms of open loop modal characteristics. Application of such requirements in design could imply a much greater reliance on analytic methods to evaluate or "demonstrate" flying qualities. In practice, however, the approach has often been to use these methods to provide a configuration which is finalized using piloted simulation. Reference 4 claims the results of this pragmatic approach have been less than successful.

The emphasis on precision tracking is compatible with the commonly used linear, small perturbation analysis techniques. However, we must remember there are large amplitude maneuvers that also require precise pilot control. Examples are target acquisition and precise heading changes. Such maneuvers require different considerations of pilot dynamics as well as analytic capability for nonlinear, coupled dynamic modes.

Flying qualities requirements on airplane dynamics based on task needs are exemplified by the limits placed on the time delay of airplane response to pilot command (Ref 3, para. 3.5.3). The time delay defined here is the effective delay - that sensed by the pilot. The pilot's ability to achieve acceptable precision tracking is quite sensitive to this delay and the numerical requirement of no more than 0.1 second delay for Level I flying qualities is based on experimental validation of values necessary to maintain a pilot rating of 3.5 or better, as the delay increases. (A Cooper-Harper rating of 3.5 or better constitutes Level I flying qualities.) Figure 3 illustrates the impact of time delay on closed loop characteristics when the pilot is included. As the figure shows, the stability reduction due to the pilot's 0.3 sec delay dominates the closed loop characteristics for $\omega_{sp} \geq 2$ rad/sec, and in effect limits the bandwidth. The open loop requirements, on the other hand, tend to minimize phase shift and extend the bandwidth.

FLYING QUALITIES CONSIDERATIONS IN DESIGN

At this point in our discussion of flying qualities, specification and pilot modeling relationships it might be well to review the general definition of flying qualities, shown below:

> Flying Qualities: Those stability and dynamic response characteristics of an airplane and its control system which impact the pilot's ability to complete some useful task or mission.

As you can see this very general definition of flying qualities involves the dynamic properties of the airplane as the pilot sees it - that is, with all augmentation and control dynamics (in both normal and failed states). The second important point is the relationship to a mission event or task. In other words, the design goal of achieving "constant flying qualities" across the Operational Envelope does not necessarily mean "constant dynamics". Finally, the purpose of the machine is to assist or enable the human operator to accomplish some job. Hence, the characteristics of the airplane should not make it an encumbrance. With this definition in mind, let us now consider how flying qualities enters the design process.

Figure 4 depicts the system design problem as a hierarchy of decision making which begins with identifying some need and determining how well it must be satisfied. The system problem becomes a design problem only if the required capability does not exist. At this point detailed requirements are needed to guide the design of a new or modified system. It is at this point that flying qualities requirements are usually considered to appear. However, availability of general task performance related requirements at the higher decision-making level could favorably impact system design decisions. A couple of general comments about this hierarchy are in order before continuing. First, note it is not a one-way path; iterations can occur at and between all levels. Second, the decision hierarchy is very general and can be repeated at each level.

To illustrate the way in which flying qualities enters the design problem, let us consider the block diagram of Figure 5. This diagram clearly shows the importance to flying qualities requirements of the interface between the human and the machine. It further illustrates a difficulty in specifying the flying qualities requirements for the airplane. That is, the pilot's perception of the vehicle's characteristics is strongly influenced by his environment. As the figure shows, this environment involves aspects of the cockpit displays, the vehicle dynamics and task requirements. The dynamic nature of this interface underscores the idea that flying qualities is not a constant in the system design. As Figure 6 depicts, definition of the requirements as well as the flying qualities is a dynamic part of the design process.

In it's earlier forms the specification could have accommodated the application of pilot modeling. Performance in the required flight test maneuvers could readily have been calculated using an appropriate form of pilot model. Note that model forms for open loop commands may not correspond to tracking models. Performance standards and pilot limitations would then have completed flying qualities requirements suitable for both analytical and flight test demonstration compliance. This is speculation and there are many arguments against that approach. It is also academic since, as we have seen, the military specification developed along the lines of open loop modal parameters. We need, therefore to consider the current and future direction of the requirements.

The implication (strong but not intended) that MIL-F-8785B forced design toward classical dominant mode characteristics has been corrected in the "C" revision. The same requirements now apply to equivalent parameters representing the overall airplane response to pilot input, not to a particular mode of the system. Figure 7, from Reference 5, illustrates the use of equivalent parameters to demonstrate compliance with requirements. In this example the actual system is modeled using a classical short period equivalent system. As can be seen, some questions remain to be answered regarding interpretation. The backup document also discusses alternate criteria including closed loop criteria. Thus, the specification applies to designs that can be represented by modal characteristics and allows others to meet alternate requirements. This approach is being continued and expanded in the latest revision effort. Drafts of the proposed MIL-Standard and Handbook are currently being distributed for critique. For the pitch axis, equivalent system parameters shall meet requirements as in MIL-F-8785C. Plus, an alternate is stated in terms of bandwidth of the (open loop) pitch attitude response. This is intended to address the closed loop piloting requirements more directly, in a form that is amenable to the design process and also flight test evaluation. We intend to continue developing the requirements in this direction.

THE USE OF PILOT DYNAMIC MODELS

Analytic methods to describe the pilot's behavior evolved because of the long recognized need to account for the pilot's effect on aircraft responses. Greater use of dynamic analysis during design has resulted in refinement of pilot modeling techniques in both the classical servo and optimal control approaches. Emphasis on the task relationship to flying qualities has created some interest in greater use of pilot modeling in defining flying qualities requirements. However, the whole premise behind current pilot modeling and its relationship to flying qualities is that closed loop tracking is the most demanding flying qualities test. With the advent of missiles, computing sights, trainable guns and improved sensors this may no longer be a valid assumption. The pilot may spend much less time tracking, and the tracking he does may require less precision. The implication of this change in the pilot's role is that the servo and optimal control tracking models will no longer be adequate by themselves. There will be a need to consider decision making and workload along with intermittent control activity. Further, identification methods based on single axis small perturbation approaches will not be appropriate. To provide a baseline for these future considerations, let's consider where we are with respect to pilot modeling methods as applied to flying qualities requirements.

Applications of pilot modeling to flying qualities criteria development have been based on the classical or optimal control tracking models, and have followed two tracks. One, use knowledge of the pilot dynamics to define vehicle characteristics that will result in acceptable closed loop tracking responses. Two, use knowledge of the pilot model parameters to predict or correlate with the pilot's subjective assessment of the airplane characteristics (i.e., Cooper-Harper Rating). Neal and Smith (Ref 6) used crossover pilot-vehicle analysis to propose criteria for pitch attitude dynamics, as shown in Figure 8. Chalk (Ref 7) used whis work as the basis for a proposed new pitch dynamics requirement for MIL-F-8785B, Figure 9. Chalk's intent was to define requirements for the augmented open loop airplane dynamics considering the pilot-aircraft closed loop properties. More recently, Hoh (Ref 8) has proposed tentative specification requirements using a definition of the vehicle frequency response bandwidth, Figure 10. His contention is that bandwidth is a fundamental parameter for correlating vehicle dynamics with pilot acceptance for some particular task. The common element in these efforts has been to specify airplane characteristics in terms of frequency response parameters.

Another approach to flying qualities criteria, specification in terms of task performance, was proposed by Onstott (Ref 9). This approach parameterized the flying qualities problem from the viewpoint that task performance was primary and airplane dynamics should be adjusted to fit an assumed pilot capability to get desirable task performance. In contrast to the previously described criteria, Onstott defined requirements in the time domain. As Figure 11 shows, Onstott defined his criterion in terms of closed loop task performance measures, rms error during target acquisition and time-on-target during a finite period of gunsight tracking. The pilot rating boundaries shown were generated empirically and appear logical in a heuristic sense. That is, small rms error and long time-on-target should warrent a good pilot rating - unless excessive workload is required to achieve the performance. Further analysis of the pilot parameters would address the workload question. Also, the data presented here are for pitch axis tracking. The more realistic multiaxis case should be investigated to fully evaluate this approach to defining criteria.

The second track in applying pilot model analysis to flying qualities was initiated by Anderson (Ref 10). His technique, called paper pilot, was based on the parameters of the servo pilot model. Subsequently, Hess (Ref 11) Schmidt (Ref 12) and Levison (Ref 13) have pursued rating prediction methods based on the optimal pilot model. Hess proposed scaling the quadratic performance index of the optimal model to get a Cooper-Harper rating prediction. Schmidt further investigated Hess' approach in an effort to develop an analytic control synthesis methodology which would specifically address flying qualities considerations. Levison proposed a rating prediction method similar to Anderson's but using elements of the optimal control model.

Specifically, he suggested rating expressions which are weighted sums of the probability of task related variables exceeding certain limits and the pilot's attention level; in other words, a combination of performance and workload factors.

Now, finally, how does all this relate to pilot identification? First, of course, is the need to validate pilot models with experimental data. Both the servo and optimal control models have been defined based on tracking experiment data. Furthermore, application of the optimal model requires specifying values for a large number of parameters for which experimental data provides baseline values. Second, experimental evaluation of flying qualities for mission-related tasks leads to the need for identifying the pilot component of closed loop responses in order to verify the airplane dynamics. Identification of the pilot's dynamics is difficult under ideal conditions because of the inherent time delays, remnant and potential time-varying behavior. In the flight test environment the task is made even more difficult by the problem of defining the input signals to the pilot. For example, standard correlation or spectral analysis methods require a known random-appearing external forcing function. Elkind (Ref 14), Wingrove (Refs 15, 16) and Heffley (Ref 17) have proposed alternative methods based on measurements normally available. Elkind compared power spectral identification methods with a method based on orthonormal functions. Wingrove showed how the pilot's time delay could be used to advantage in estimating pilot dynamics when the random forcing function is not available. Heffley and Jewell adapted Wingrave's idea to devise an "on-line" method for generating coefficients of an assumed pilot equation from a closed loop task.

A different viewpoint toward pilot identification than that expressed by most people arises from the use of pilot opinion rating and associated comments to both validate flying qualities criteria and evaluate airplane flying qualities. This reliance on the pilot's subjective rating stimulated the efforts mentioned previously to predict pilot ratings using model parameters and measures of task performance. Smith and Torgerson (Ref 18) considered the Cooper-Harper rating to be a vector and investigated methods to identify the factors and states the pilot includes in his rating vector. Their preliminary results indicated about six parameters dominate his evaluation but more data are needed to statistically isolate them from a larger set. Identification of these states would provide another means of correlating pilot acceptance with airplane characteristics.

One of the most difficult elements of the pilot's characteristics to quantify is workload. The difficulty, of course, is because of the indirect way workload enters the problem. Reference 19 discusses several concepts of workload as well as a number of metrics. However, no clear conclusions or recommendations emerge. The importance of workload to rating prediction methods, task performance oriented criteria and the increasing complexity of the pilot's job as a system manager all point to a need for a better capability to account for this aspect of the pilot's characteristics.

CONCLUSIONS

We have shown that flying qualities was recognized very early in the history of manned flight to be a dominant consideration in the dynamic pilot-airplane interface. As a formal entity, flying qualities initially spun off from stability and control, and performance considerations. Many requirements in the military flying qualities specification are still presented in terms of classical airplane modal dynamic parameters. As a consequence of the impression that these modal requirements would not apply to highly augmented aircraft, the specification has been ignored or circumvented in some cases. The reasoning in these cases has been that high authority stability and flight control systems would alleviate any potential problem (e.g., prevent high angle-of-attack problems by limiting control stick authority). Efforts have been made in the current modifications to the military specification to clarify and improve the applicability of requirements to anticipated highly augmented aircraft, (e.g., equivalent system discussion and alternate "bandwidth" requirements).

It is recognized, however, that mission requirements will play an even greater role in future military aircraft design. As these requirements become more stringent, the effective integration of aerodynamics, performance and control will be essential to meeting them. Specification of open loop airplane characteristics based on implied or indirect closed loop requirements will not be adequate. It appears knowledge of the pilot's dynamic and performance characteristics is necessary to quantitatively relate airplane dynamic properties to many task-oriented design requirements. Many such requirements are closed loop in that they involve the pilot as an active controller. Others involve the pilot intermittently as a controller or decision maker. The Flying Qualities Group has an effort planned to develop closed loop criteria to consider future piloting requirements more directly. Both the development of these requirements and their application will require identification and analysis of the pilot's dynamics and workload.

REFERENCES

1. McRuer, D. T. and E. S. Krendel, "Mathematical Models of Human Pilot Behavior", AGARD-ograph, AGARD-AG-188, January 1974.

2. A'Harrah, R. C. and R. J. Woodcock, "The Military Flying Qualities Specification, A Help or Hindrance to Good Fighter Design". Presented at the AGARD Flight Mechanics Panel Symposium, Florence, Italy, October 1981.

3. Anon., "Flying Qualities of Piloted Airplanes". Military Specification, MIL-F-8785C, 5 November 1980.

4. Smith, R. E., "Evaluating the Flying Qualities of Today's Fighter Aircraft". Presented at the AGARD Flight Mechanics Panel Symposium, Florence, Italy, October 1981.

5. Moorhouse, D. J. and R. J. Woodcock, "Background Information and User Guide for MIL-F-8785C, Military Specification - Flying Qualities of Piloted Airplanes". AFWAL-TR-81-3109, Air Force Wright Aeronautical Laboratories, WPAFB, Ohio, September 1981.

6. Neal, T. P. and R. E. Smith, "An In-Flight Investigation to Develop Control System Design Criteria for Fighter Airplanes". AFFDL-TR-70-74, Vols I & II, Air Force Flight Dynamics Laboratory, W-PAFB, Ohio, December 1970.

7. Chalk, C. R., et al., "Revisions to MIL-F-8785B(ASG) Proposed by Cornell Aeronautical Laboratory Under Contract F33615-71-C-1254". AFFDL-TR-72-41, Air Force Flight Dynamics Laboratory, W-PAFB, Ohio, April 1973.

8. Hoh, R. H., et al., "Development of Handling Qualities Criteria for Aircraft with Independent Control of Six Degrees of Freedom". AFWAL-TR-81-3027, Air Force Wright Aeronautical Laboratories, W-PAFB, Ohio, April 1981.

9. Onstott, E. D. and W. H. Faulkner, "Prediction, Evaluation, and Specification of Closed Loop and Multiaxis Flying Qualities". AFFDL-TR-78-3, Air Force Flight Dynamics Laboratory, W-PAFB, Ohio, February 1978.

10. Anderson, R. O., "A New Approach to the Specification and Evaluation of Flying Qualities". AFFDL-TR-69-120, Air Force Flight Dynamics Laboratory, W-PAFB, Ohio, June 1970.

11. Hess, R. A., "Prediction of Pilot Opinion Ratings Using an Optimal Pilot Model". In Human Factors, Vol 19, No. 5, 1977, pp. 459-475.

12. Schmidt, D. K., "On the Use of the OCM's Quadratic Objective Function as a Pilot Rating Metric". In Proceedings of the Seventeenth Annual Conference on Manual Control, UCLA, June 16-18, 1981, pp. 305-312.

REFERENCES CONCLUDED

13. Levinson, W. H., "A Model-based Technique for Predicting Pilot Opinion Ratings for Large Commercial Transports". NASA CR-3257, Langley Research Center, April 1980.

14. Elkind, J. 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, Air Force Flight Dynamics Laboratory, WPAFB, Ohio, March 1964.

15. Wingrove, R. C. and F. G. Edwards, "A Technique for Identifying Pilot Describing Functions from Routine Flight-test Records". NASA-TN D-5127, Ames Research Center, May 1969.

16. Wingrove, R. C., "Comparison of Methods for Identifying Pilot Describing Functions from Closed-loop Operating Records". NASA TN D-6235, Ames Research Center, March 1971.

17. Heffley, R. K. and W. F. Jewell, "Development of a CTOL Piloting Technique Measurement Scheme for a Real-time Simulator Environment". NASA CR-152294, Ames Research Center, July 1979.

18. Smith, R. H. and W. S. Torgerson, "A Multivariate Approach to Handling Qualities Rating Scale Development". Final Report. Contract F49620-79-C-0158, AFOSR, Bolling AFB, D. C., Janaury 1980.

19. Hartman, B. O. and R. E. McKenzie, editors, "Survey of Methods to Assess Workload". AGARD-ograph, AGARD-AG-246, NATO, AGARD Aerospace Medical Panel, August 1979.

FIGURES

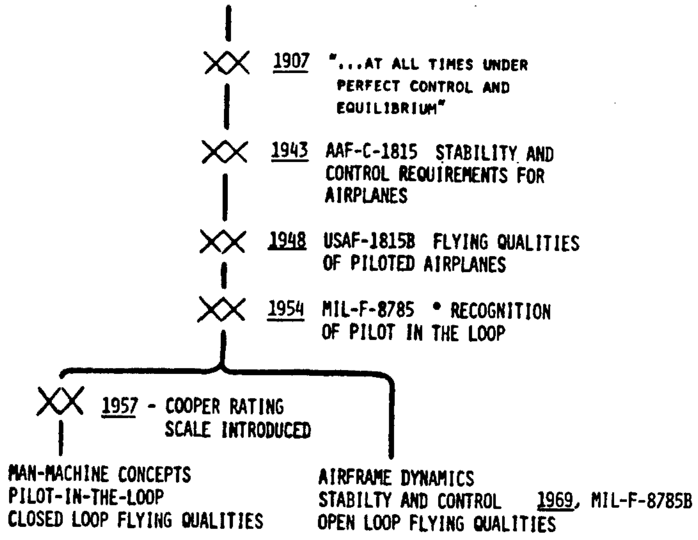


FIGURE 1. HISTORY OF FLYING QUALITIES

E-Io. <u>LONGITUDINAL CONTROL:</u>

•
∘
∘

E-Io(I)(A). <u>LONGITUDINAL CONTROL AT HIGH SPEEDS.</u> - ALL CLASS III AIRPLANES SHALL DEMONSTRATE THEIR ABILITY TO RECOVER FROM HIGH MACH-NUMBER DIVES. THIS TEST SHALL BE MADE BY DIVING VERTICALLY 15,000 FEET FROM SERVICE CEILING AT WHICH TIME PULL OUT WILL BE STARTED EXCEPT IN NO CASE WILL PULL OUT BE STARTED BELOW 15,000 FEET. AIRPLANES IN THIS CLASS SHALL DEMONSTRATE THEIR ABILITY TO RECOVER FROM SUCH DIVES AT NORMAL ACCELERATIONS OF AT LEAST 3G BY THE APPLICATION OF A STICK FORCE NOT EXCEEDING 150 POUNDS OR THROUGH THE USE OF ANY OTHER RECOVERY DEVICE.

•
∘
∘

FIGURE 2. EXAMPLE OF EARLY FLYING QUALITIES REQUIREMENT

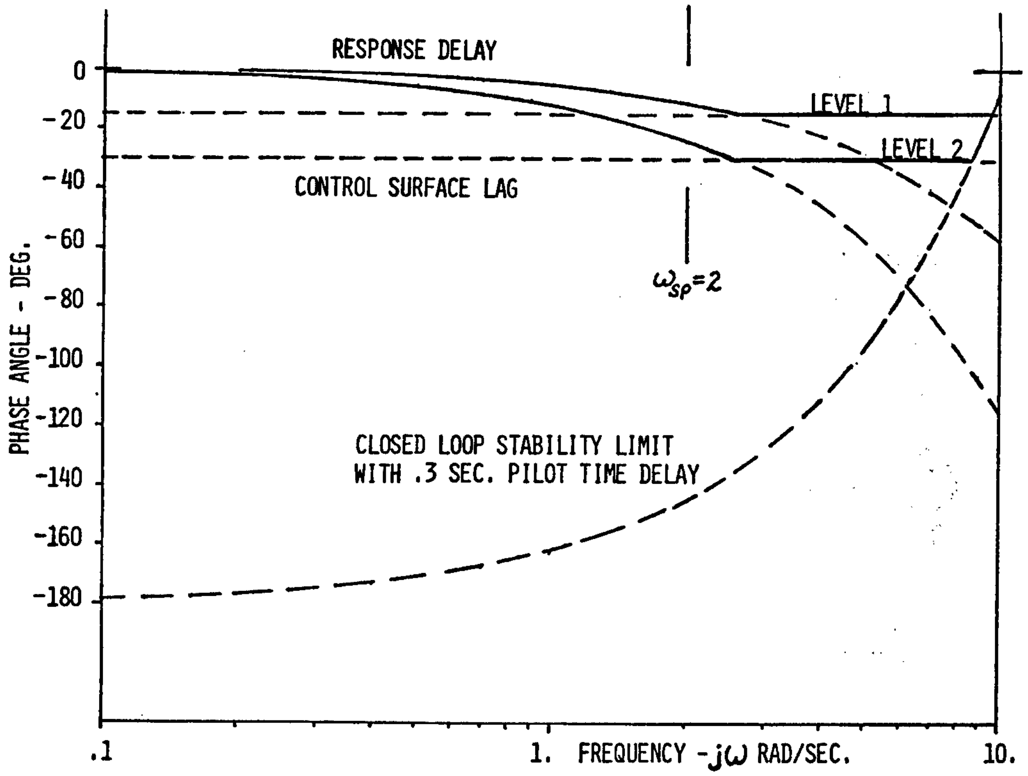


FIGURE 3. EFFECT OF TIME DELAY ON CLOSED LOOP STABILITY

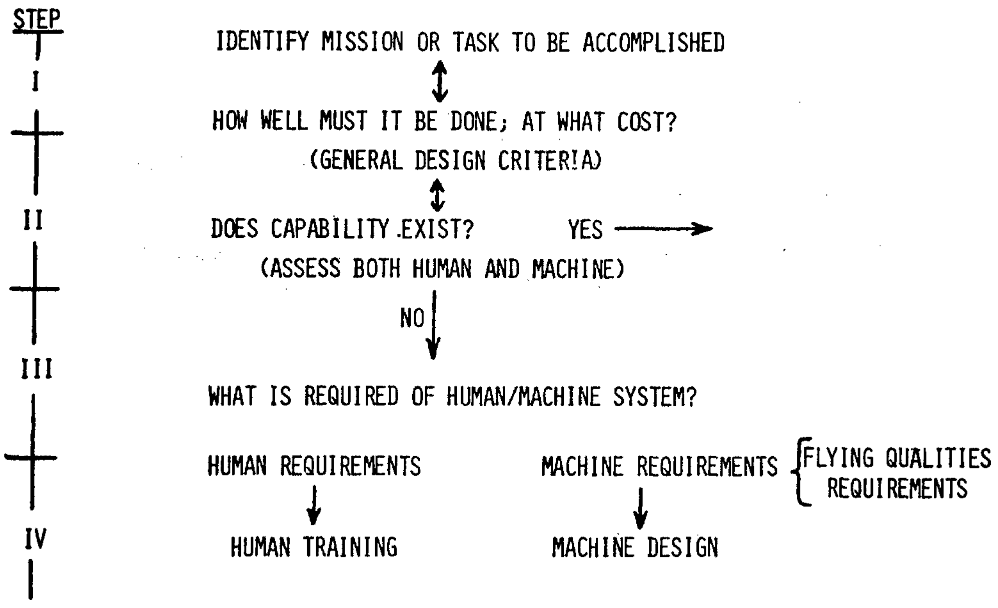


FIGURE 4. THE SYSTEM DESIGN PROBLEM - A DECISION HIERARCHY

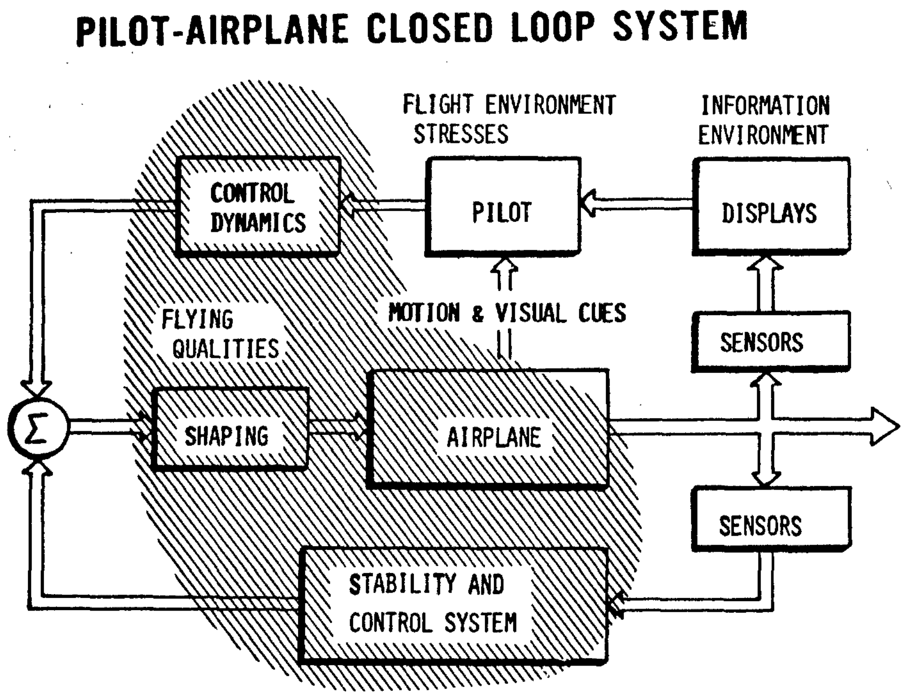


FIGURE 5. PILOT-AIRPLANE FLYING QUALITIES INTERFACE

CLOSED LOOP FLYING QUALITIES ANALYSIS

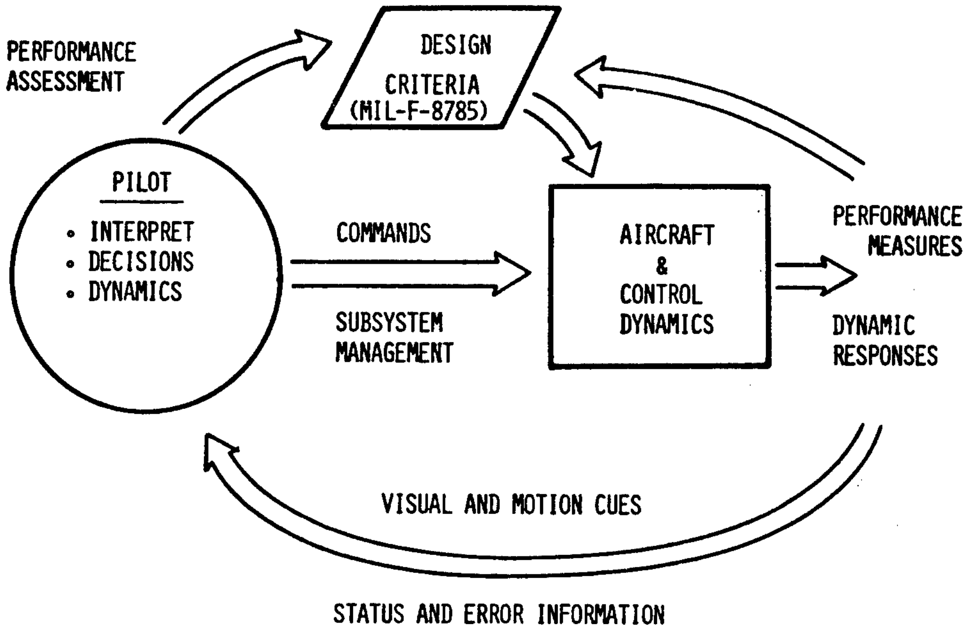


FIGURE 6. DYNAMIC INTERACTION OF CRITERIA IN DESIGN PROCESS

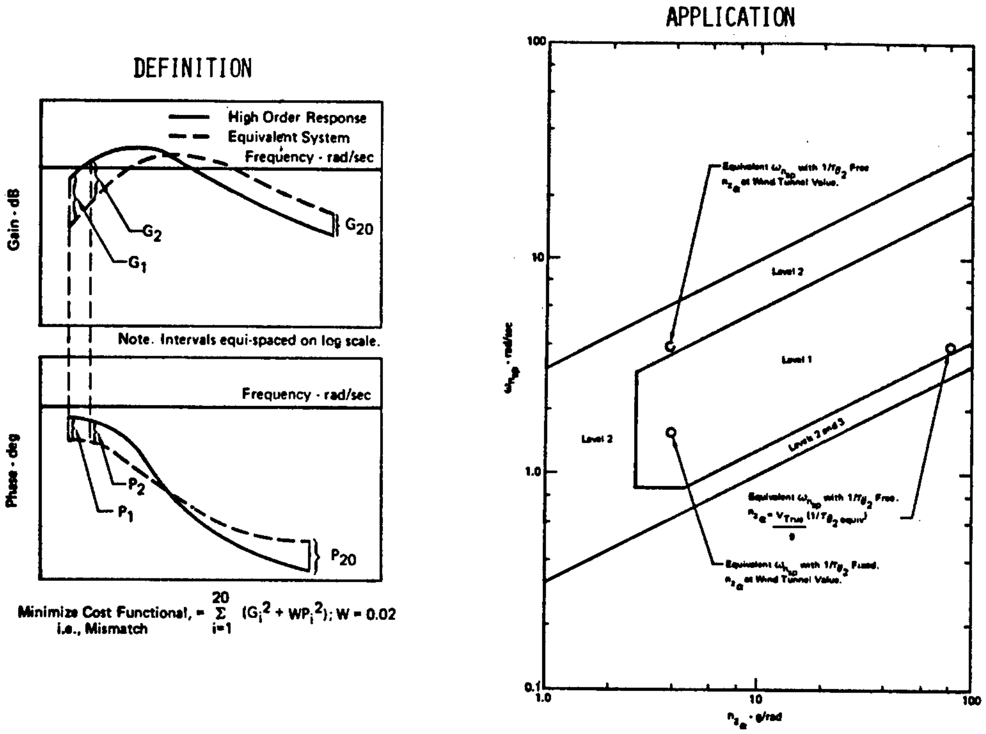


FIGURE 7. DEMONSTRATING COMPLIANCE WITH EQUIVALENT PARAMETERS

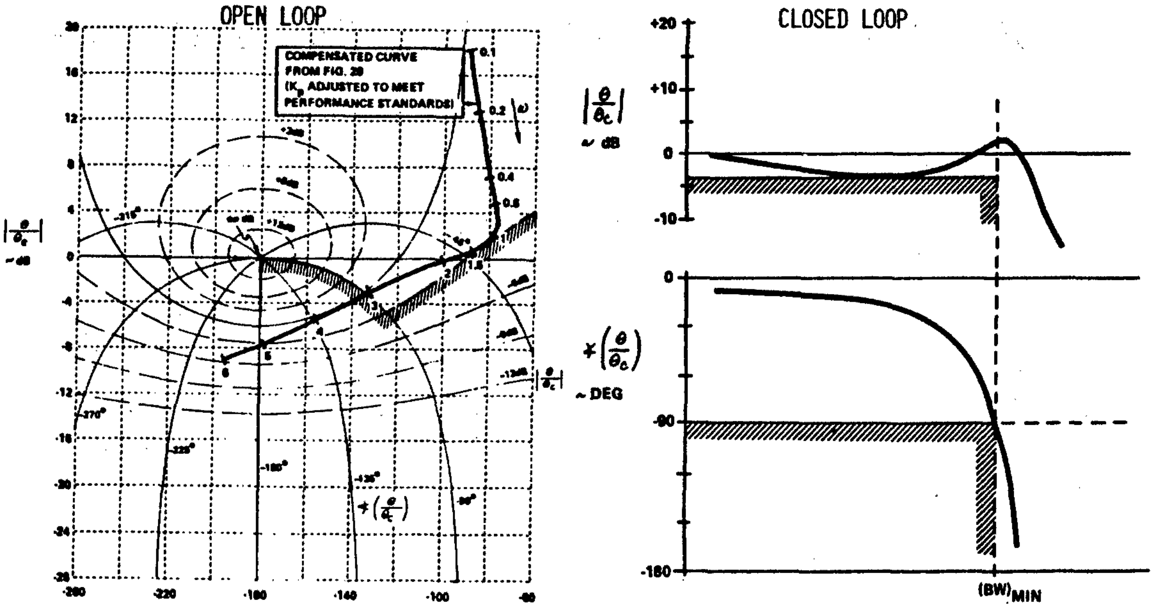
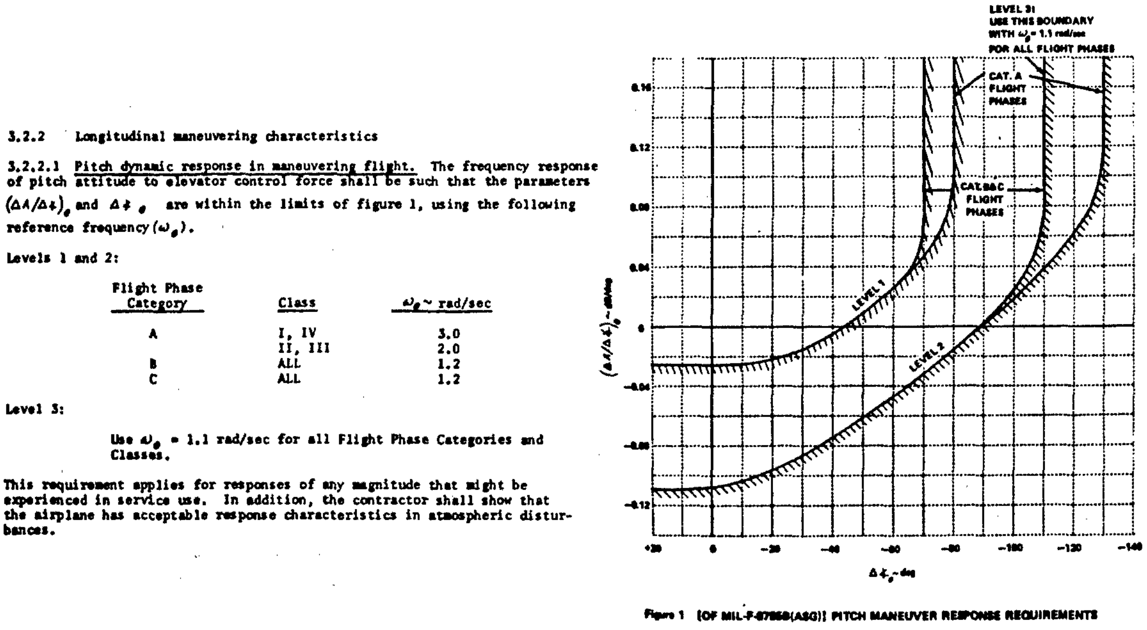


FIGURE 8. NEAL-SMITH PITCH DYNAMICS CRITERIA



3.2.2 Longitudinal maneuvering characteristics

3.2.2.1 Pitch dynamic response in maneuvering flight. The frequency response of pitch attitude to elevator control force shall be such that the parameters $(\Delta A/\Delta \phi)_e$ and $\Delta \phi_e$ are within the limits of figure 1, using the following reference frequency (ω_e).

Levels 1 and 2:

Flight Phase Category	Class	ω_e ~ rad/sec
A	I, IV	3.0
	II, III	2.0
B	ALL	1.2
C	ALL	1.2

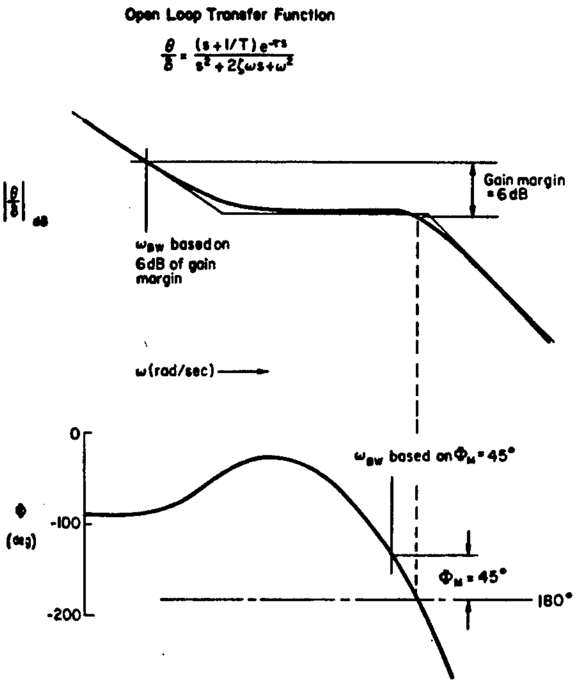
Level 3:

Use ω_e = 1.1 rad/sec for all Flight Phase Categories and Classes.

This requirement applies for responses of any magnitude that might be experienced in service use. In addition, the contractor shall show that the airplane has acceptable response characteristics in atmospheric disturbances.

Figure 1 [OF MIL-F-8785B(ASG)] PITCH MANEUVER RESPONSE REQUIREMENTS

FIGURE 9. CHALK'S PROPOSED PITCH DYNAMICS REQUIREMENT

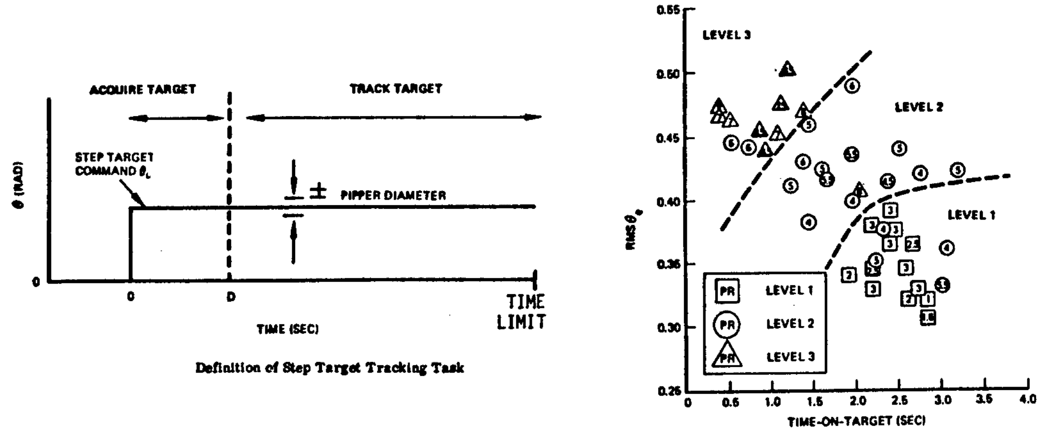


TENTATIVE BANDWIDTH LIMITATIONS

TASK	REQUIRED BANDWIDTH (rad/sec)	
	LEVEL 1	LEVEL 2
Tracking (CAT A) Air-to-air gunnery Strafing Photo Dive bombing	1.25	0.60
Path deviation (CAT C) Formation Air-to-air refueling Approach	0.30	0.12
Short final and landing path response ("CAT D")	$(\dot{H}_P - 3)/10^*$	?

$^*\dot{H}_P$ = sink rate in ft/sec on visual or instrument glide slope

FIGURE 10. HOH'S PROPOSED BANDWIDTH CRITERION



Definition of Step Target Tracking Task

Pilot Ratings as Functions of rms θ_e and Time-On-Target

FIGURE 11. ONSTOTT'S PITCH TARGET TRACKING CRITERION

DEVELOPMENT OF CONTROLLER REQUIREMENTS FOR UNCOUPLED AIRCRAFT MOTION

J. Hodgkinson, K. D. Citurs (McDonnell Aircraft)
R. H. Hoh, (Systems Technology, Inc.)
T. J. Cord, (AFWAL)

The use of uncoupled, six-degree-of-freedom (6 DOF) motion is rapidly becoming state of the art in terms of necessary flight control laws and aerodynamic capability. McDonnell Aircraft and Systems Technology Inc. are under contract to AFWAL to develop design criteria for cockpit control devices for 6 DOF motion which will assure compatibility among the pilot, the control device(s) and aircraft response. This will allow efficient implementation of the 6 DOF control capability and improved mission performance. The effort includes a study of existing data followed by a ground-based simulation. In the simulation planning, appropriate methods for estimating the impact of 6 DOF capability on pilot activity, will be defined. The estimate will be in terms of pilot workload, controller activity and display requirements. It will be used to explore the tradeoffs of manual vs automatic control for those aircraft classes and tasks where 6 DOF capability is of benefit.

The presentation will be a progress report briefly describing the study of existing data, and outlining the simulation plan. Emphasis will be placed on the plans for pilot workload estimation.

A FLIGHT RESEARCH VIEWPOINT

Donald T. Berry and Terrence W. Rezek

My comments will be brief since much of what I had in mind can be summed up by endorsing what has already been said, particularly in regard to the papers by Col. Milan, Col. O'Donnel, and Ron Anderson.

However, from the point of view of a flying qualities flight research engineer and a human factors engineer with a strong involvement in the pilot-vehicle interface, my remarks can be categorized under three headings:

1. Needs for Flight Research
2. Needs of Flight Research
3. Needs of Flight Test

Not all the items I will touch on are pilot dynamics or workload items in themselves, but they all are strongly related in that pilot dynamics and workload measurements would greatly enhance the research I will discuss. With respect to the first item, needs for Flight Research, I will discuss three subtopics:

1a. In-Flight measures of pilot strategy and gain

1b. Simulation validation for critical tasks

1c. Parametric data on pilot controllers

In regard to item 1a:

In the fall of 1980 a workshop on pilot induced oscilations was held at Dryden. One of the main conclusions of the workshops was that existing pilot in the loop analysis techniques were highly successful in identifying generally PIO prone or PIO resistant configurations. However, to apply these techniques as accurate design tools, more data was needed on pilot gain and control strategy as a function of vehicle class, dynamics and task. Flight measurement of these characteristics is needed because pilot gain and strategy is often different in a ground simulator than it is in flight. Which leads us to item 1b, "Simulator Validation for Critical Tasks."

An area of concern is the difference between real world flight and simulation with respect to pilot behavior in response to the stress of critical situations. For most situations, particularly involving routine flight management tasks, flight and ground simulator results compare very well. However, for many critical tasks, such as final phases of landing, abrupt collision avoidance, aerial refueling etc., significant differences are observed between pilot technique and behavior in flight as compared to simulator results. An example of this is illustrated by figure 1, which compares ground and flight results for pitch PIO tendencies of an advanced aerospace vehicle. The ground simulator was the Ames FSAA (Flight Simulator for Advanced Aircraft) and the airborne simulator was the Air Force/Calspan TIFS (Total In-Flight Simulation). The horizontal axis is the PIO rating scale number as-signed by the pilots. A rating of 1 indicates no PIO tendency, and a rating of 6, an uncontrollable PIO. It can be seen that there is as significant difference in PIO tendencies between ground and airborne results. Because of the heavy reliance on ground simulation in the design and development process, many advanced prototype aircraft have been found to have serious deficiencies in the human factors and displays area. The cost of redesign, modification, and flight test to alleviate such problems is extremely high. It would be highly productive then, to determine the limits of applicability of ground simulators for these situations, and strive to obtain a basic understanding of the differences between flight and ground results so that they might be eliminated or ameliorated in the future.

A basic tool in this process would certainly be pilot dynamics and workload measurements for identical scenarios in flight and in ground based simulators.

Pilot controller is the third area recommended for flight research. Despite some excellent work in this field, pilot controllers have been designed on an ad hoc basis, and nobody seems really satisfied with the results. A systematic investigation of controller characteristics for a reasonable range of parameters does not exist. This is especially true for sidestick controllers. And, because of the sensitivity of controllers to g, stress, pilot cues, motion, etc., a flight validated data base is needed. And, of course, good measures of pilot workload are needed to implement such research.

The second category to be discussed are the needs of flight research. That is, those items that can contribute to the increased effectiveness and efficiency of human factors flight research. Three subtopics will be discussed here, namely:

2a. Non-intrusive measurement techniques and compact lightweight instrumentation

2b. Remotely Piloted Vehicle (RPV) utilization in human factors research

2c. More communication and interaction between flying qualities engineers and human factors engineers

With respect to item 2a we are often faced with a dilemma. In order to determine what a pilot is doing in a particular task, we often have to intrude with measurements or procedures that interfere with the task at hand. Consequently the task we want to study is altered! A need exists for non-intrusive means of making the desired measures that do not alter the task at hand or encumber the pilot with unrealistic procedures. There is also a need for research equipment compatible with the confines of a fighter cockpit. A case in point is the oculometer. Presently, these systems are bulky and heavy and require elaborate calibration procedures. This negates their routine use in flight test except in large aircraft under limited circumstances. Efforts are underway to develop more practical oculometer systems and these are to be applauded. It is hoped that similar effort can be applied to other equipment. Many advanced aircraft are flown routinely at the Dryden Flight Research Facility. If they could be routinely instrumented to measure pilot scan patterns, workload, etc., a large data base for a variety of tasks and vehicle configurations could be gathered at low cost.

Remotely piloted research vehicles, (RPRVs) are being used to conduct highrisk flight research or operations. It has been observed that RPRV missions generate high pilot gains during critical flight conditions. The motivation and stress levels produced during RPRV operations approach, and can occasionally surpass, those encountered during onboard operations. Such levels can not be duplicated on even the most expensive simulator.

In addition, input/output devices (individual displays, control mechanisms, etc.) or entire systems may be changed in the ground cockpit with greater ease and less expense and personal risk than in a fully human rated aircraft. The computers that reside with RPRV systems easily accommodate a variety of signal manipulations for information transfer (displays), control modification, or data acquisition. RPRVs then, offer a cost effective way of bridging the gap between the laboratory and the real world. Figure 2 illustrates the RPRV technique as applied to the highly maneuverable aircraft technology (HiMAT) vehicle.

The interaction between allied disciplines such as flying qualities and human factors is becoming more active, but a case can be made for some crosstraining to achieve a greater appreciation of each other's methodology. For example, human factors engineers tend to use large numbers of subjects and rely on statistical inference. Flying qualities engineers tend to use a small number of highly expert well calibrated subjects and rely on pilot commentary and subjective rating scales. Combinations of these techniques may be the best way to approach pilot modeling and workload measurement.

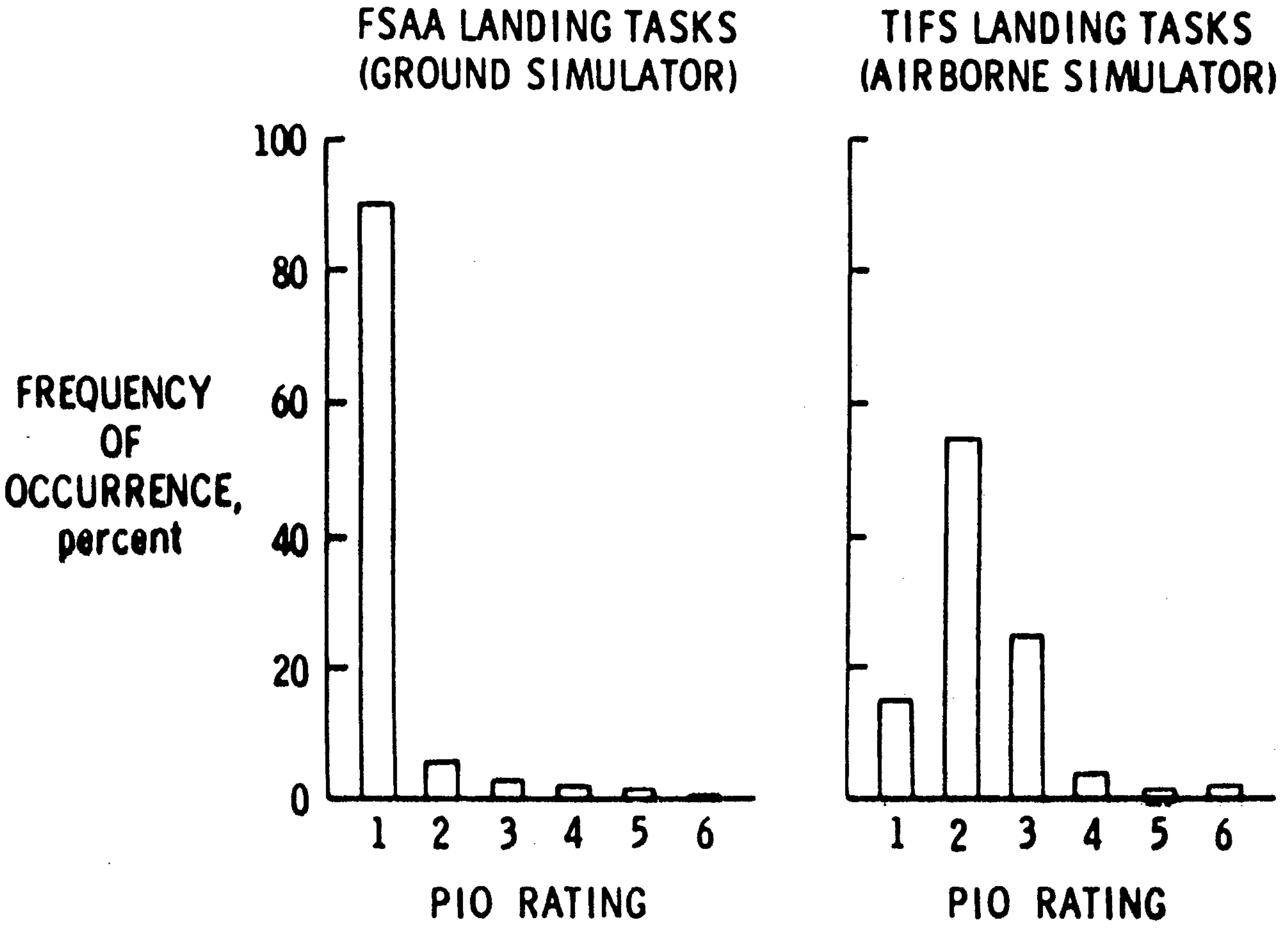
Needs of Flight Test

The third and final area I would like to touch on relates to the contributions human factors and display technology can make to flight testing itself. This consideration may be somewhat peripheral to the main theme of this workshop, but is sufficiently related to warrant consideration.

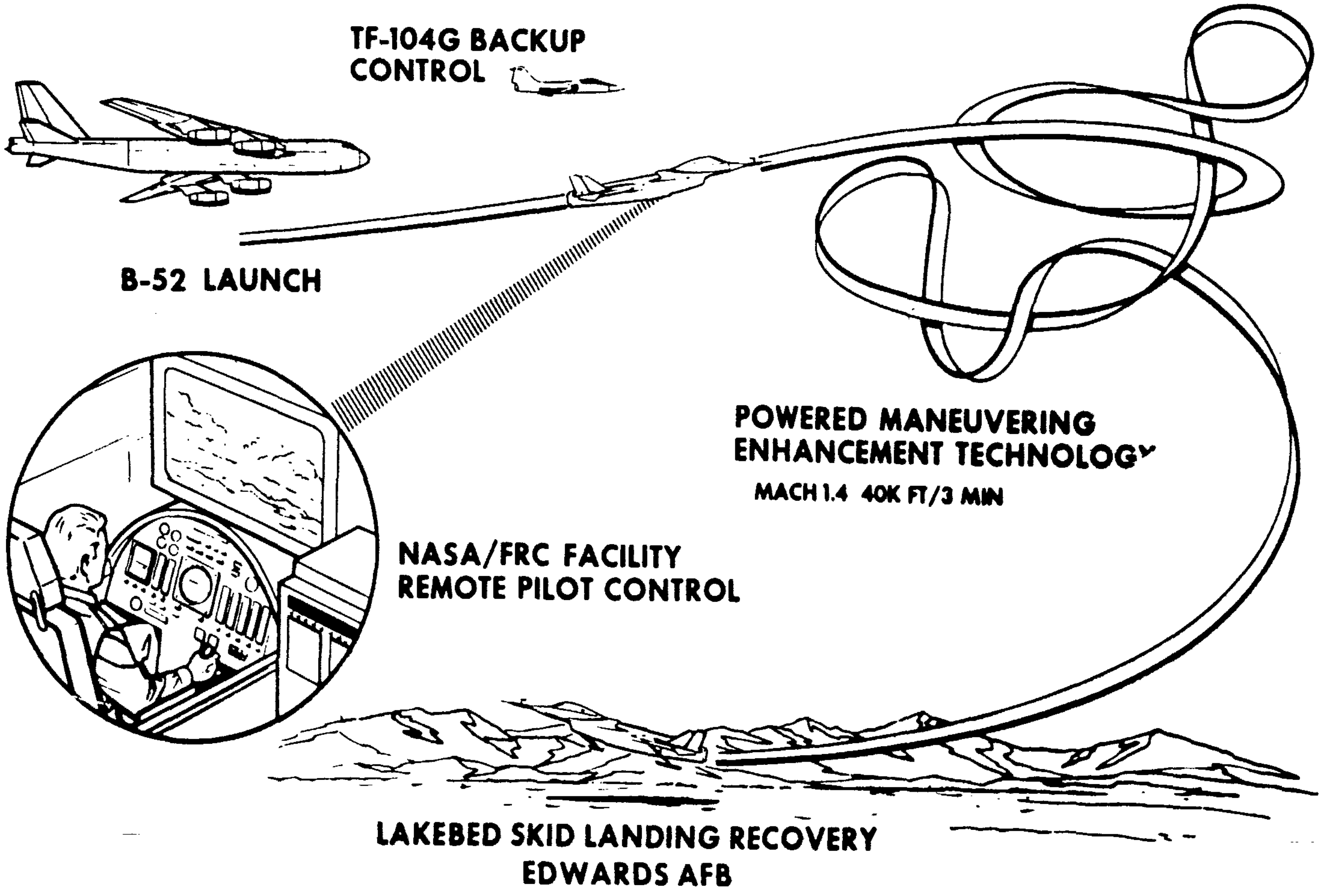
For the most part, pilot dynamics and workload is thought of in the context of operational missions such as air combat, approach and landing etc. We are all familair with the complex tasks associated with various mission phases. However, a flight test is a mission also, and although it has many tasks in common with other missions, in many cases the requirements are unique. For example, in NASA flight research, it is often required to precisely match wind tunnel test conditions and flight conditions so that accurate correlations can be made. Extremely close tolerances on several variables may be required to be held simultaneously for extended periods of time. For example, a typical test may require Mach number to be held ±.01, and angle of attack and angle of sideslip to be held within ±.25°, during a level turn for at least one minute. Such a task is extremely difficult with conventional displays, requiring significant repeats to obtain good quality data, and subsequent increases in expensive flight time. With a special purpose integrated display, however, the task was performed with relative ease and required precision. Nevertheless, design of the display was a time consuming "cut and try on the simulator" process. Better information about pilot dynamics and workload would greatly facilitate the design process for this application.

In conclusion then, we have seen that there is a need for flight research in pilot dynamics and workload measurement. Also, pilot dynamics and workload measurements can contribute to more effective human factors flight research as well as to more efficient flight testing in general.

PITCH PIO RATING RESULTS



HiMAT OPERATIONAL CONCEPT



NASA Flight Research Center
Edwards, California

SECTION II

TAILORING FLIGHT TESTS TO IDENTIFY PILOT WORKLOAD

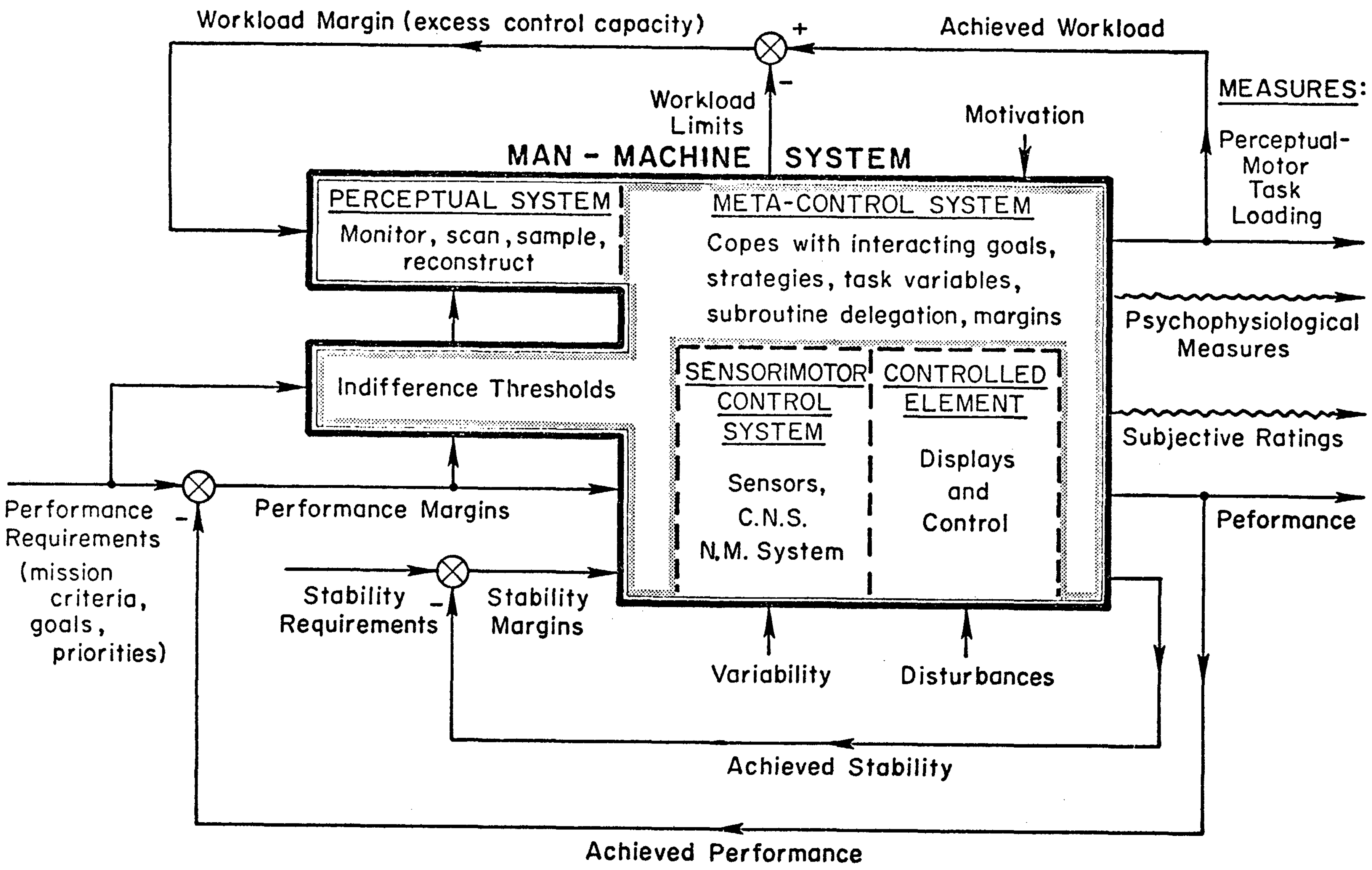
Abstract For EAFB Workload Workshop, 19-21 January 1981

MEASURING AIRCREW WORKLOAD:
PROBLEMS, PROGRESS, AND PROMISES

Henry R. Jex
Systems Technology, Inc.
Hawthorne, California 90250

An overview of the problems of defining, quantifying, and measuring mental workload during aircrew tasks is given based on our work in the areas of aircraft handling qualities, pilot model measurement and prediction, multi-display scanning and psychophysiological correlates of workload. The continued promise and problems with psychopysiological measures is assessed and the importance of some new multidimensional workload rating techniques is emphasized. The lack of unifying theoretical approach is identified as the main impediment to progress, and an approach is suggested, that can handle both continuous and discrete task loads. A review is given of some new workload measurement concepts such as Non-Invasive Pilot Identification Program, the "imbedded surrogate" auxiliary task method, and the measurement of workload margin via the Cross-Coupled-Instability Task (CCIT).

THE PROBLEM : BLOCK DIAGRAM SHOWING INTERACTIONS AMONG MAN-MACHINE STABILITY, PERFORMANCE, AND WORKLOAD



CRITERIA FOR WORKLOAD MEASURES

1. RELEVANT:

   TO PROBLEM & SOLUTION

2. SENSITIVE:

   MONOTONIC W.R.T SUBJECTIVE WL. HIGH "TEST-POWER" WRT WL VARIABLES. INSENSITIVE TO OTHER VARIABLES

3. CONCORDANT:

   UNIVERSAL EFFECTS IN TARGET POPULATION

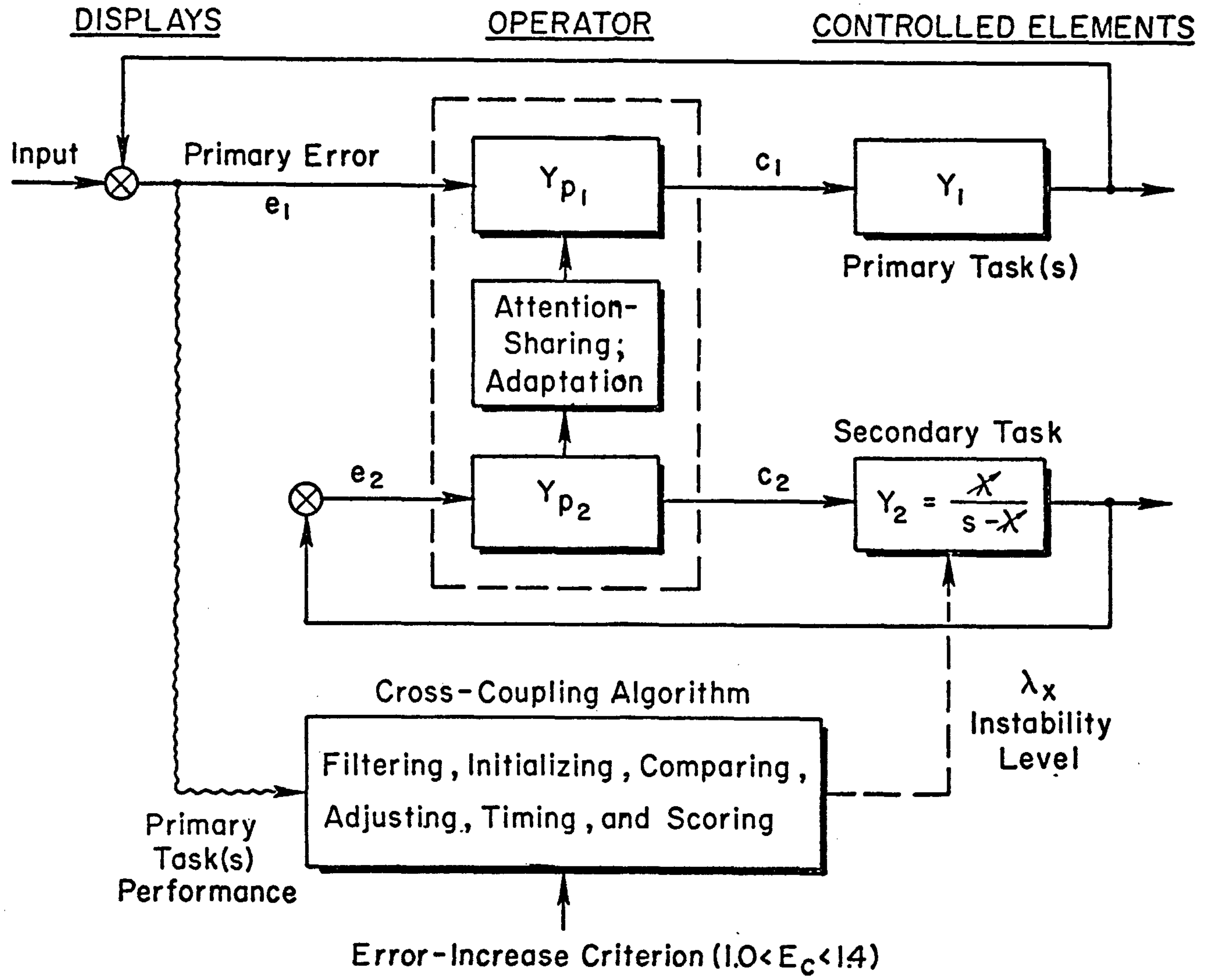
4. RELIABLE:

   TEST-RETEST
   "DIFFERENTIAL STABILITY" WRT PRACTICE
   VALIDATED; WITH NORMS

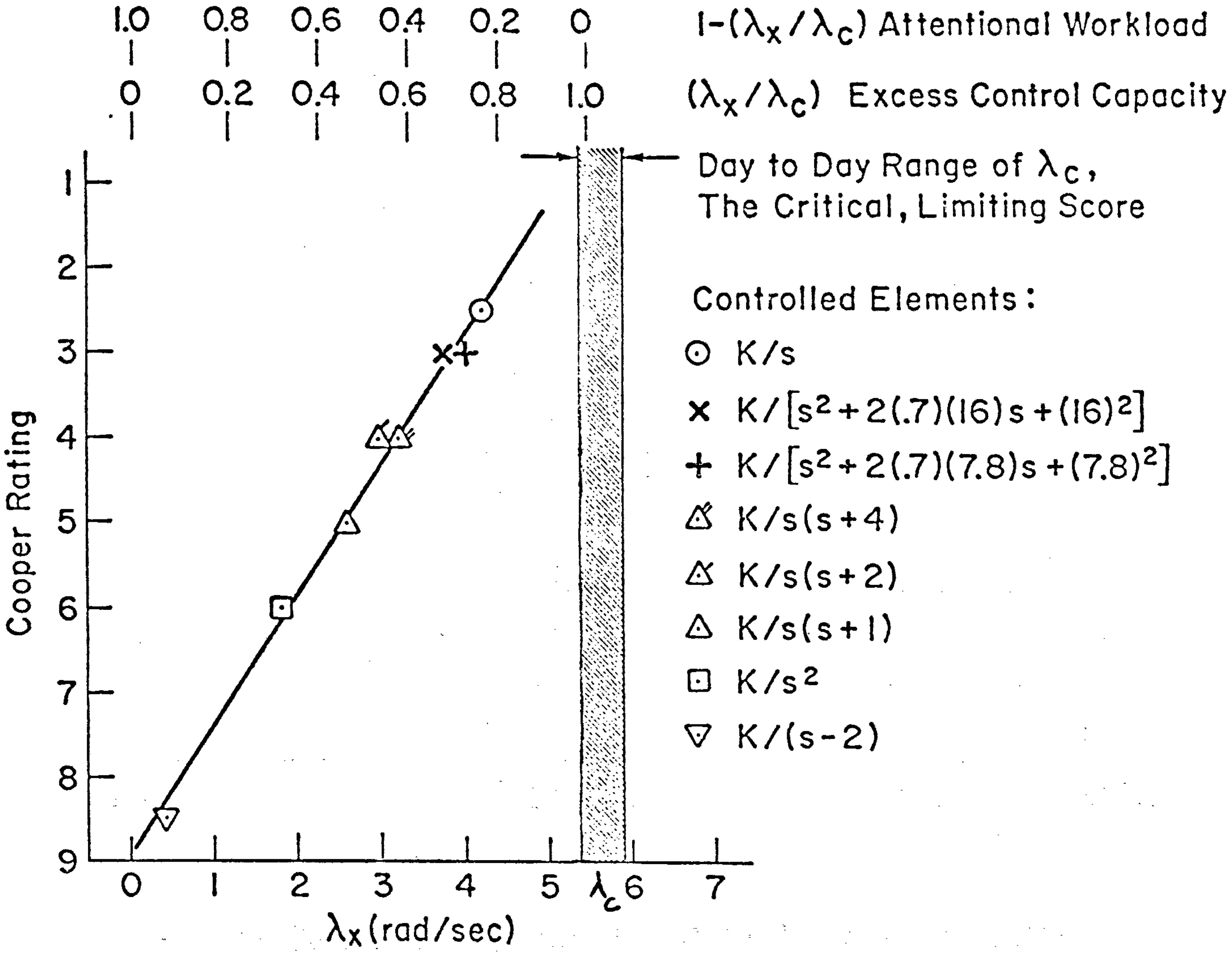
5. CONVENIENT:

   PORTABLE
   EASY TO LEARN

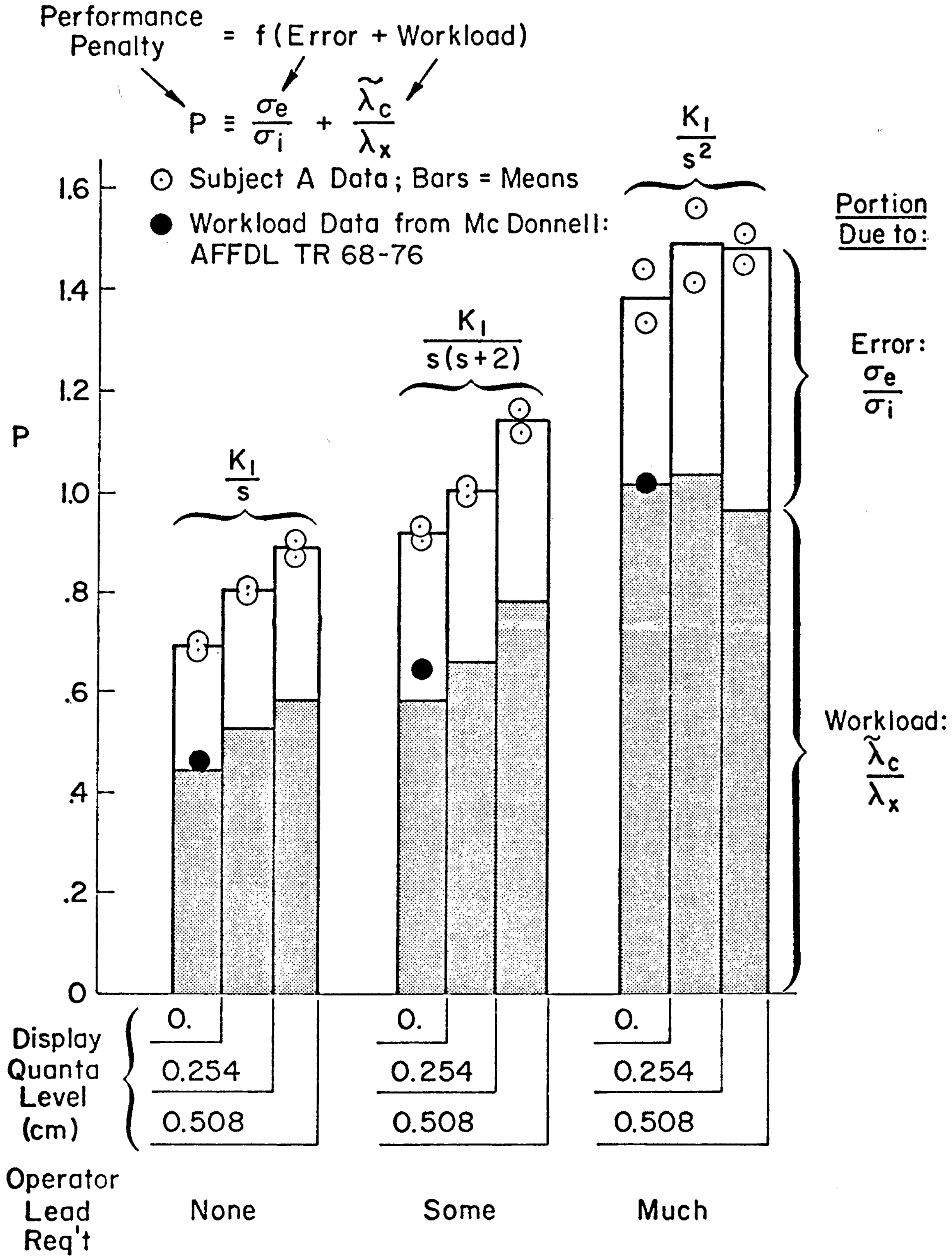
ELEMENTS OF THE CROSS-COUPLED INSTABILITY TASK (CCIT)



CROSS ADAPTIVE MEASURE OF EXCESS CONTROL CAPACITY FOR SEVERAL EXAMPLES OF PRIMARY CONTROLLED ELEMENTS



TYPICAL APPLICATION OF ADAPTIVE-WORKLOAD TESTING



EVALUATION OF A PILOT WORKLOAD ASSESSMENT DEVICE
TO TEST ALTERNATE DISPLAY FORMATS AND
CONTROL HANDLING QUALITIES

by
Samuel G. Schiflett, Ph.D.
Naval Air Test Center
Patuxent River, Maryland

Paul M. Linton
Naval Air Development Center
Warminster, Pennsylvania

Ronald J. Spicuzza
Systems Research Laboratories
Dayton, Ohio

SUMMARY

This in-flight research project evaluated the utility of a Workload Assessment Device (WAD) to measure pilot workload for approach and landing tasks under simulated instrument meteorological conditions, alternate HUD formats and control stability variations. The flight tests were conducted in an NT-33A research aircraft, extensively modified for the U. S. Air Force and U. S. Navy by the Display Evaluation Flight Test program. The hardware, software, and test procedures associated with the WAD functioned efficiently with only minor discrepancies and minimum pilot distraction. The project established the feasibility of using an item-recognition task as a measure of sensory-response loading and reserve information processing capacity while flying precision approaches. In a descriptive statistical treatment of the data, the results indicate an appreciable increase in reaction time and errors with degraded handling qualities as compared to ground baseline measures and good handling qualities. The preliminary findings also reveal consistent trends toward the availability of more mental reserve capacity when flying predominantly pictorial/symbolic HUD configurations as compared to conventional HUD formats with scales and alphanumerics. It is recommended that further evaluations be conducted to establish the efficacy of utilizing the WAD to measure mental workload in a wide variety of aircrew tasks.

INTRODUCTION

BACKGROUND

New developments in cockpit display designs and integrated weapons system avionics have significantly altered the role of the pilot from that of a skilled, manual control operator to an executive manager of an integrated weapons system. Emphasis on psychomotor control has been augmented by an interest in more cognitive skills represented by such functions as short-term memory, information processing, and decision making. Few measurement techniques exist which are able to provide an objective, reliable, and valid estimate of the subtle differences in workload introduced by these new systems. To date, methodology for objectively quantifying workload has not been effectively applied to the flight test and evaluation of aircrew systems (references 1 and 2).

This project introduced a novel approach to the traditional manner of measuring pilot workload. Aircrew workloads are typically measured by subjective assessment rating scales which are based on pilot opinions that relate operational task demands to system response characteristics, e.g., Cooper-Harper Handling Qualities Rating Scale. The new approach applied in this project is an item-recognition task first identified by Sternberg (reference 3) and modified by the U. S. Air Force (reference 4) to measure the reserve capacity of the pilot. The approach assumes that an upper bound exists on the ability of the pilot to gather and process information. As the pilot's workload increases on the primary task, i.e., flying the aircraft, reserve capacity for processing secondary information decreases until a point of overload is reached by the pilot. At this point, the information processing demands of the task exceed the pilot's total workload capacity and is manifested by degradation in performance (i.e., increase in errors and response times) on the secondary item-recognition task.

The theoretical formulation of the item-recognition task, as proposed by Sternberg (figure 1), has several attractive features which make it ideally suited for evaluating the source of increase in task-loading in aircraft test environments. The theory assumes a least-squares, linear regression fit of the data where the intercept represents the input/output component and the slope depicts the mental information processing component of the item-recognition task. If, for example, the sensory-response mode (i.e., input/output), is response overloaded the theoretical expectation is a change in the y-intercept of the regression line with no change in slope. Conversely, if the source of task-loading was one which affected the pilot's mental information processing capabilities (e.g., short-term memory overload), the expectation is a change in the slope of the curve without a corresponding change in the intercept value. Either result would be a decrease in the pilot's reserve capacity for processing information.

The use of the item-recognition task to assess primary task workload is not a new concept in aircrew flight simulation studies (references 5 and 6). However, the uniqueness of its application in this project is that a Workload Assessment Device (WAD) that generates and controls the secondary item-recognition task was designed, fabricated, and installed in a NT-33A research aircraft to measure and analyze the pilot's reserve workload capacity for the Display Evaluation Flight Test (DEFT) program as reported in reference 7.

PURPOSE

The purpose of this project was to evaluate the utility of the WAD to measure pilot workload for approach and landing tasks under simulated Instrument Meteorological Conditions (IMC's) for alternate HUD formats and aircraft control stability variations.

DESCRIPTION OF AIRCRAFT/EQUIPMENT

The NT-33A variable stability aircraft is an extensively-modified, T-33 jet trainer. The elevator, aileron, and rudder controls in the front cockpit were disconnected from their respective control surfaces and connected to separate servo-mechanisms that comprise an "artificial feel" system. In addition, the elevator, aileron, and rudder control surfaces were connected to individual servos which were driven by a number of different electrical inputs. This arrangement, through a response-feedback system, allowed the normal T-33 stability derivatives to be augmented to the extent that the handling qualities of the hypothetical research configurations could be simulated. A more comprehensive description of the NT-33A can be found in reference 8.

The DEFT program also provided a fully software-programmable display system to complement the variable stability features of the host-modified NT-33 Aircraft. Relative to the aircraft configuration, the DEFT system provided the capability of changing display formats and changing the algorithms and dynamics of the display driving signals. The display system consisted of a HUD, two digital computers, a magnetic tape system, INS sensors to augment the existing aircraft sensors, and a display repeater and mode control unit for the aft cockpit.

The software programs provided an in-flight choice of two uniquely different display configurations for use in the approach and landing phases of flight. These displays were of a conventional HUD format (figure 2) and the predominantly symbolic Klopfstein format (figure 3). As depicted in the figures, the conventional display used a HUD format with a flight path ladder, scales, and alphanumeric readouts of various flight parameters. The Klopfstein display, however, is predominantly symbolic/pictorial depicting the horizon, and artifical runway overlaying the actual runway, and other flight guidance symbols.

METHOD

After several practice sessions and prior to the start of the evaluation flights, a baseline measurement was obtained on the item-recognition task. Each pilot was given the item-recognition task for each memory set size while sitting in the cockpit of the aircraft stationed on the ground. The task required the pilot to memorize sets of one, two, or four letters, i.e., A, RJ, ZPNW. The pilot was then instructed, prior to testing with each memory set, which set of letters would be presented for memory recall. The prememorized letters (positive) or other letters (negative) were presented on the HUD one at a time every 7 sec. The positive and negative letters were presented individually with a .5-probability of occurrence. Each letter appeared on the HUD one at a time until the pilot responded or 5 secs. elapsed. The pilot responded to a letter presentation by pressing one of two designated buttons on the control stick. One button indicated that the letter was a member of the prememorized set (positive) and the other indicating it was not a member of the prememorized set (negative). Positive letters never appeared as negative letters and the same positive letter sets were used throughout the test. A total of 30 letters, 15 positive and 15 negative, was presented for each memory set for the baseline conditions.

The same procedures were used in flight as during the baseline test conditions with the exception that the pilot was flying the aircraft while performing the secondary task. An additional experimental control allowed one approach per handling quality/display format combination to be flown without any letter presentations to evaluate the impact of the secondary task on the primary task of flying the aircraft.

The reaction times and response errors were collected and analyzed by the WAD controller and ground-based analysis system. After each response, the reaction time was measured from the onset of a letter to the physical response of pressing the correct button. The reaction times for both the positive and negative letters were stored on cassette tapes. The reaction times for the correct responses were then averaged and plotted as a function of the memory set sizes. The response errors were coded, tabulated, and categorized by type of error and frequency of occurrence. A response was considered an error if the pilot pressed the wrong key (reversal error), responded correctly but after 1,500 msec (out-of-bound error), or did not respond before 5 sec (time-out error).

The basic flight scenario for each approach and touch-and-go was as follows. The Evaluation Pilot (EP) was given control of the aircraft by the Safety Pilot (SP) with the desired display-aircraft handling quality combination. The EP then flew on instruments while using an orange filter over the windscreen and a blue visor attached to the helmet to simulate IMC.[1] After intercepting the glide slope, the EP descended to 1,800 meter MSL to intercept the localizer at 8 nmi. At this point, the SP turned on the digital recorder and the WAD controller which were used to record the primary flight measures and the secondary task measures, respectively. The EP proceeded to fly the glide slope and the localizer to perform the approach. The outer marker was at approximately 4 nmiles. At 200 meter AGL and approximately 1/2 nmi from the runway threshold, the EP "broke out" (i.e., he lifted the blue visor) and flew visually for the remainder of the low approach (7 meters AGL). If conditions permitted (fuel state, crosswind, etc.), the EP then performed the touch-and-go landing, minimizing the sink rate on touchdown to less than 1 meter/sec. The touchdown point was a 170-meter zone, 500 meters from runway threshold. After liftoff and at approximately 70-meter AGL, the SP turned off the WAD controller and the digital recorder. After four approaches, the SP assumed control of the aircraft, then changed the pitch handling quality to the next desired setting and again released control of the aircraft to the EP.

After each block of four approaches was completed under the same pitch handling quality, the EP and SP rated the approach and flare/landing segments of the flight profile using the Cooper-Harper pilot rating scale. Additional commentary data were gathered from the EP and SP throughout the flight tests by use of an audio tape recorder, e.g., comments on degree of air turbulence.

The WAD consists of two basic units: the airborne controller and the ground-based analysis center. The controller is configured for installation in the front avionics bay of the NT-33A research aircraft. The unit provides the electronics, power supply, software, interfaces to the HUD and the aircraft intercom, rear cockpit initialization switches, control stick response switches, and data recording system necessary to perform a complete series of item-recognition experiments. In addition, the controller can operate as a stand alone laboratory system capable of performing the same tasks as when airborne. The ground-based data analysis center is used to initialize several software options of the controller and to reduce and analyze response time data. A description of the functional capabilities of the hardware and software is discussed in appendix A. A detailed description of the complete WAD system is contained in reference 9.

SCOPE

Each pilot flew two evaluation flights using the conventional HUD format and two with the Klopfstein format. During each evaluation flight, a pilot performed eight approaches terminating in either a low approach or touch-and-go landing for a total of 32 approaches per pilot. One-half of the approaches for each flight were made using "good" handling qualities, the other half were made using either "fair" or "poor" handling qualities. The handling qualities were manipulated by changing the pitch response (150 msec or 200 msec time delay) of the aircraft after every four approaches. The response of the roll and yaw axes was held constant throughout the tests.

RESULTS AND DISCUSSION

GENERAL

The test and evaluation paradigm used in this project was a repeated measures design in which type of display format (conventional versus Klopfstein), flight handling quality (good versus poor), and secondary task difficulty (memory set sizes, 0, 1, 2, and 4) were fractionally combined to form 16 different conditions. It was planned that the two EP's would be exposed to each of the 16 conditions twice. However, each EP was able to complete all combinations of the test conditions only once. Out of a total of eight 1.5-hr evaluation flights, a complete set of secondary task data was analyzed for four flights only.

The results showed that the general procedures established for the conduct of the evaluation flight tests of the WAD were acceptable to the pilots. The in-flight test procedures provided the EP's and SP's with reliable guidelines for efficient and safe crew coordination during successful approaches and during incidents of all equipment malfunctions. Pilot comments aided in the investigation of the most salient characteristics of the item-recognition task including the selection, location, and timing of the letters as presented on the HUD. A thorough testing of the WAD procedures during the project resulted in only minor software changes and hardware replacements and clearly established the feasibility of using the item-recognition task for in-flight tests.

PRIMARY FLIGHT MEASURES

The primary flight measurement data taken from the digital recorder were divided into two defined categories of approach and flare/landing. Because of the length and complexity of the analyses of the primary flight measurement data, the results were published under separate cover in reference 10.

The summary results of these analyses indicate that the primary flight performance parameters and Cooper-Harper ratings showed a general inconsistency between displays and handling qualities during the approach and flare/landing phases of the flight task.

[1] Overlaying the two complementing colors produced a perceptual environment similar to night IMC when the pilot attempted to view the external world.

Lack of systematic differences in the primary flight measures and Cooper-Harper ratings suggests that pilot performance remained the same for all conditions. That is, no significant differences were found in the primary flight measures between display formats, handling qualities, or memory set sizes. These findings indicate that the pilots compensated for the increased task difficulty by maintaining primary flight performance at an acceptable level. However, this pilot compensation was not without cost. A loss of information processing reserve capacity can be clearly shown from the results of the secondary task measures.

SECONDARY TASK MEASURES

Secondary task measures consisted of reaction times in which slopes and intercepts were calculated after solving linear regression equations for each set of data. Secondary task errors for the item recognition task were calculated for all incorrect responses, late responses, and no responses.

REGRESSION EQUATIONS (REACTION TIMES)

The reaction times associated with each correct response were averaged for the complete flight profile for each m-set size (letters 1, 2, or 4), handling quality (good or poor), and display format (conventional or Klopfstein). Linear regression equations were then calculated to indicate the slope and intercept of the plotted data as shown in figures 4 and 5. The data reveal that both the intercept and slope of the curves for each pilot increased from baseline conditions when the handling qualities were degraded. The results indicate that the WAD is sensitive to the increased sensory/response and mental processing requirements imposed by the addition of a secondary task and to the level of difficulty of that task. For example, the largest intercept and slope changes occurred between each subject's relative baseline and poor handling quality condition.

A closer examination of the data reveals that the differences in the magnitude of change in the slopes were consistently larger for the conventional HUD format than the pictorial Klopfstein HUD format under either good or poor handling qualities. This trend, relative to each subject's shift in slope magnitude, suggests that more mental reserve capacity was available to process information while flying the Klopfstein display than the conventional HUD format and while good handling qualities independent of the type of display format used.

Reviewing the resulting changes in intercepts revealed a similar trend with regard to the handling quality parameter. The average increase in the magnitude of change in intercept was less for conditions of good handling qualities than for poor handling qualities. However, with regard to the display variable, the trend was reversed from that observed for the changes in slope; i.e., the average intercept value changed less for the conventional format than for the Klopfstein. Assuming the observed trends would persist in a larger data sample, the results indicated that degrading the handling qualities had a consistent effect on the input/output stages of the item-recognition task, whereas the effect of the display format variable on the input/output stages of the task was subject to inconsistent individual differences. The lack of consistent trends in the changes in intercept relative to the display variable may be due to: (1) individual differences in establishing a time-error tradeoff,[2] (2) locations of the letter in relationship to differences in eye scan patterns, and/or (3) different strategies of memory recall.

These results suggest that degrading handling qualities had a consistent and predominant effect of reducing the pilot's reserve capacity for all three stages of the information-processing, secondary task. Changing the display formats appeared to yield similar results but are subject to the influences of individual differences with regard to the mental component of the information processing task.

The reader is reminded that these data reveal only trends and were gathered from a sample of two pilots. Additional flight data are required with a larger pilot sample and more replications of test conditions before definitive conclusions can be made concerning the reliability of the results of these measures. A further discussion of the reliability of the item-recognition task that questions the day-by-day stability of the slope and intercept is found in reference 11.

PERCENT ERRORS

The WAD provided an accumulative record of the number of errors, sequence of occurrence, type of error, and reaction time associated with each error for both positive and negative letters. The combined percent of secondary task errors for both pilots is shown in figure 6. The error data show that as the difficulty of the secondary task was increased, i.e., as the m-set size increased, a corresponding decrease in response accuracy was observed which supports the expectation of increased error rate under conditions of task overloading.

The increases found in secondary task response errors under conditions of poor handling qualities for both display formats are consistent with the results of the slope and intercept reaction time data with regard to the influence of degraded handling qualities.

[2] The EP's were only instructed to respond as quickly and accurately as possible to the secondary task while flying a precision approach and landing.

That is, under test conditions producing a reduction in reserve capacity, a corresponding increase in response errors occurred.

In contrast, the reaction time data indicated that the type of display format differentially influenced both the input/output and mental stages of the information processing tasks, whereas response error data showed a consistently higher degree of response accuracy under conditions of the pictorial Klopfstein display format.

To further explore these results, the total percent errors were classified into type of error for each handling quality and display format. The secondary task errors reveal that the total percent errors were evenly distributed between incorrect responses (reversals), late responses (out-of-bounds), and no responses (time-outs). However when the total percent errors are differentiated between display type and handling quality, it clearly shows that three times as many reversal errors were committed by the EP's flying the conventional HUD than the Klopfstein display format. Degrading the handling qualities increased the percentage of time-out errors for the EP's flying with the conventional display and increased the out-of-bounds under the Klopfstein display format. Since it was assumed that a time-out error would reflect a greater decrement in reserve capacity than an out-of-bounds error, these results would imply that the EP's had less reserve capacity while flying under the conventional HUD and degraded handling qualities than the Klopfstein display format.

In summary, the percent of secondary task errors increased whenever the memory set size increased, the handling qualities were degraded, and the task was performed in flight under the conventional display format conditions. Poor handling qualities primarily induced errors of delay or no response while the type of display affected mainly the accuracy (correctness) of response.

CONCLUSIONS

The hardware, software, and test procedures associated with the Workload Assessment Device (WAD) functioned efficiently with only minor discrepancies and minimum pilot distraction.

The project established the feasibility and sensitivity of using a secondary item-recognition task as a measure of sensory/response loading and reserve information processing capacity while flying precision instrument meteorological conditions approaches.

The pilots showed an appreciable increase in reaction time and percentage of errors on the secondary task flown under poor handling qualities as compared to good handling qualities and ground baseline conditions.

The WAD revealed that the pilots had less secondary task errors, more mental reserve capacity, but longer reaction times attributed to sensory/response delays while flying with pictorial/symbolic HUD configurations (Klopfstein) than conventional HUD formats.

REFERENCES

1. Schiflett, S. G., U. S. Naval Air Test Center, Patuxent River, Maryland. Operator Workload: An Annotated Bibliography. December 1976, SY-257R-76.

2. Wierwille, W. W., Williges, R. C., and Schiflett, S. G., Aircrew Workload Assessment Techniques. In B. O. Hartman and R. E. McKenzie (Eds.) Survey of Methods to Assess Workload. AGARD-AG-246, August 1979, 19-53.

3. Sternberg, S., High Speed Scanning in Human Memory, Science, 1966, 153, 652-654.

4. Spicuzza, R. J., Pingus, A. R., and O'Donnell, R. D., Systems Research Laboratories, Dayton, Ohio. Development of Performance Assessment Methodology for the Digital Avionics Information System. 1974.

5. O'Donnell, R. D., Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio. Secondary Task Assessment of Cognitive Workload in Alternate Cockpit Configurations. 1975. AMRL-TR-75-49, (AGARD-CPP-181).

6. Crawford, B. M., Pearson, W. H., and Hoffmann, M., Wright-Patterson Air Force Base, Ohio. Multi-Purpose Digital Switching and Flight Control Workload. 1978. AMRL-TR-78-43, (AGARD-AG-246).

7. Schiflett, S. G., Naval Air Test Center, Patuxent River, Maryland. Evaluation of a Pilot Workload Assessment Device to Test Alternate Display Formats and Control Handling Qualifier. 1980, SY-33R-80.

8. Huber, R. W. and Parrag, M. L., Calspan Corporation, Buffalo, New York. Flight Manual Supplement for the United States Air Force/Calspan NT-33A 51-4120 Variable Stability Airplane. June 1979. T.O. 1T-33A-1 Supplement (Revision B).

9. Spicuzza, R. J., Systems Research Laboratories, Dayton, Ohio. Workload Assessment Device Model WAD 8085 Manuals: Volume I - Installation/Operation, Volume II - Hardware/Software Description and Theory of Operation. 1979.

10. Monagan, S. J. and Smith, R. E., Calspan. Display Evaluation Flight Test. May 1980, 6645-F-2.

11. Carter, R. C., Kennedy, R. S., Bittner, Jr., A. C., and Krause, M., Naval Aerospace Medical Research Laboratory Detachment, New Orleans, Louisiana. Item Recognition as a Performance Evaluation Test for Environmental Research.

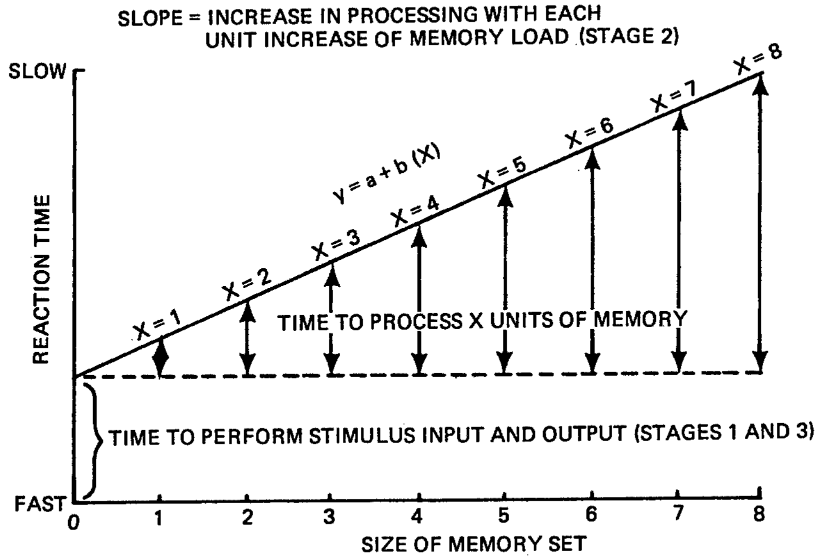


Figure 1 Theoretical Components of the Item Recognition Task Proposed by Sternberg

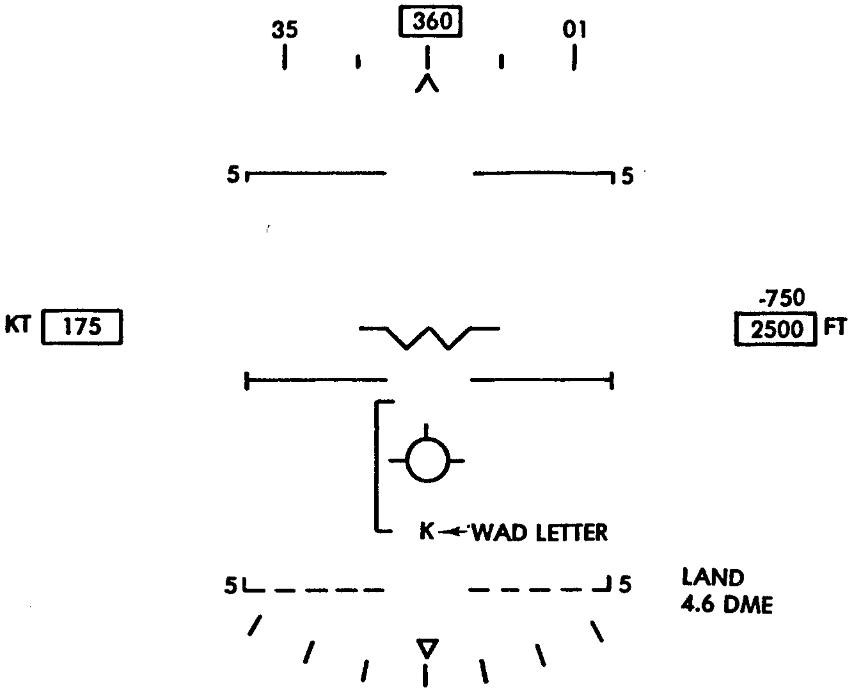


Figure 2 Conventional HUD Format

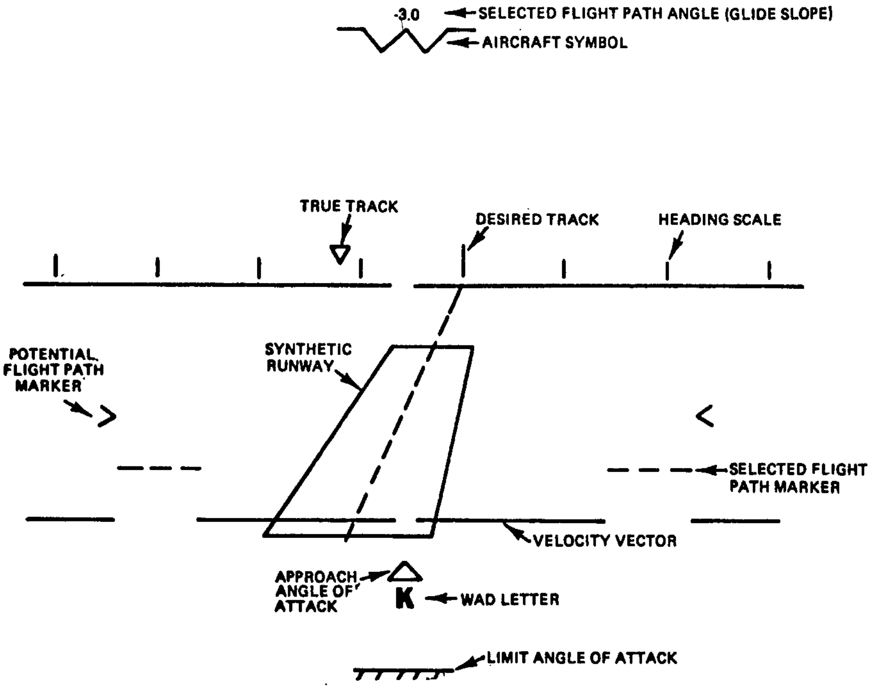


Figure 3 Klopfstein HUD Format

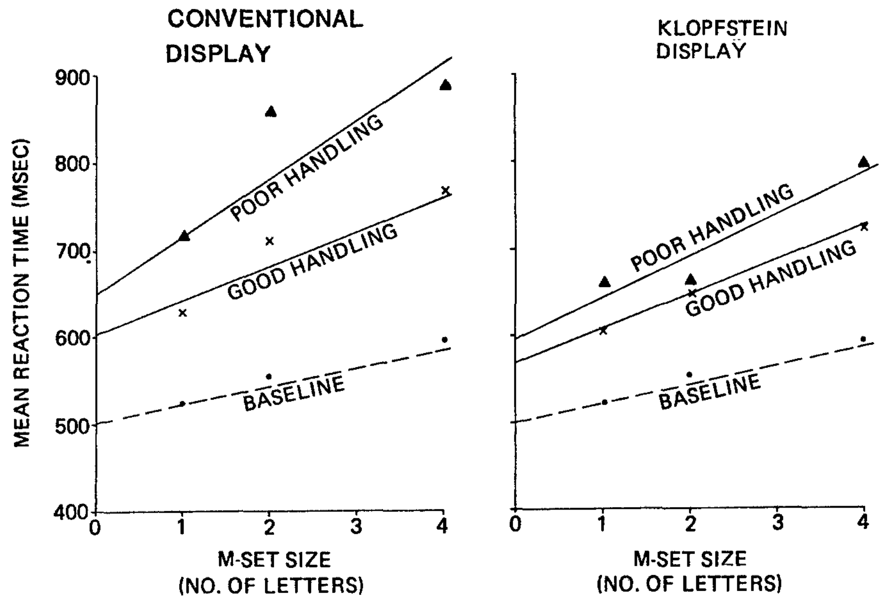


Figure 4 Linear Regression Lines for Evaluation Pilot 1

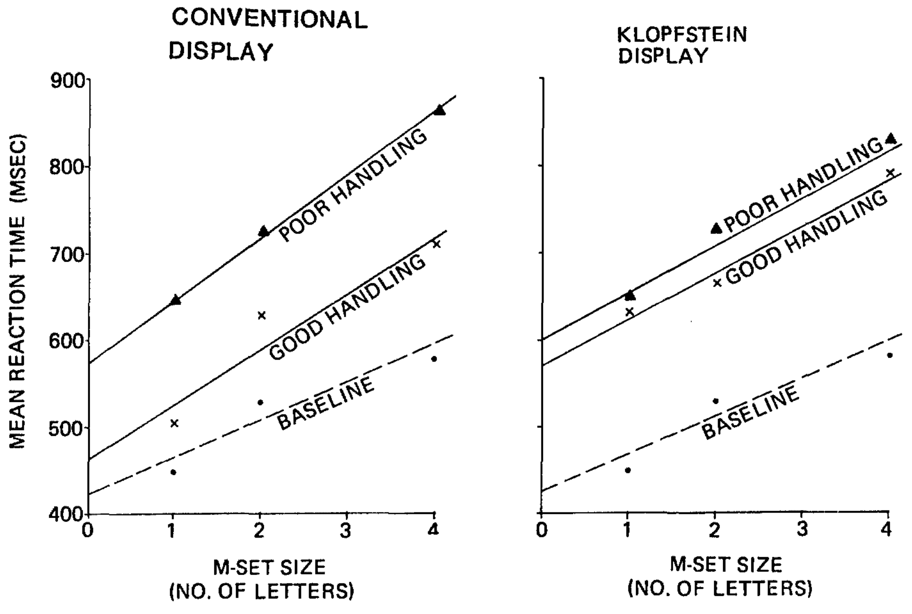


Figure 5 Linear Regression Lines for Evaluation Pilot 2

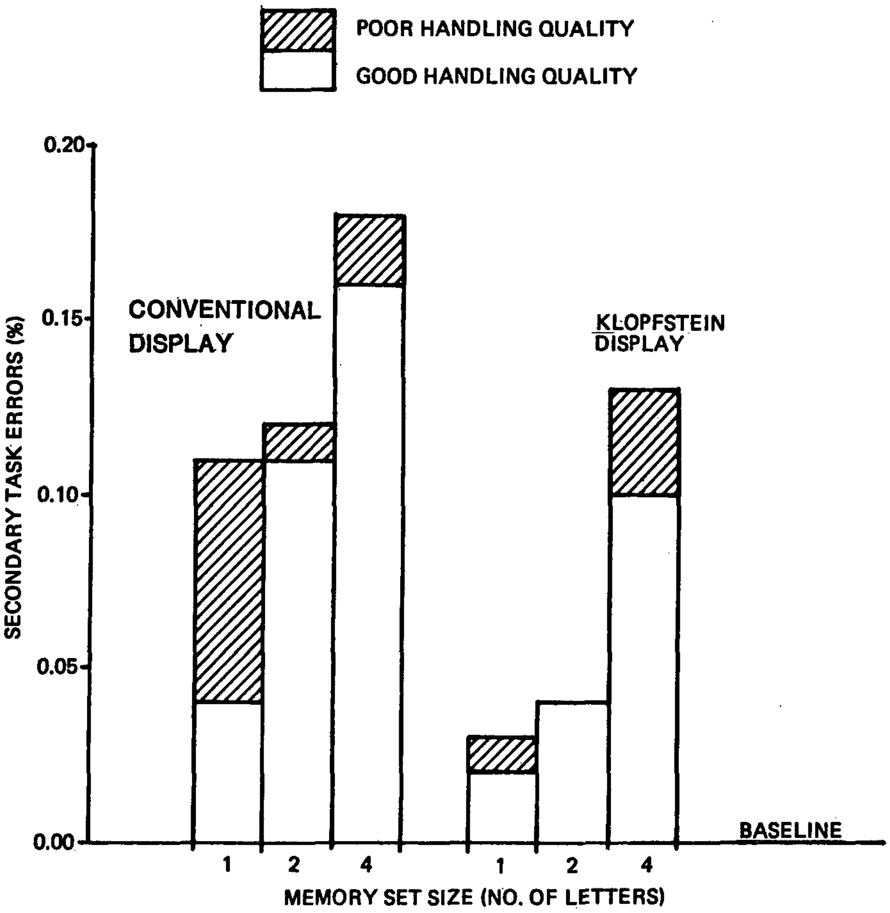


Figure 6 Mean Percent Secondary Task Error for Memory Set Size (Number of Letters) by Display Format and Handling Quality

WORKLOAD ASSESSMENT DEVICE
SYSTEM DESCRIPTION

TASK

The Workload Assessment Device (WAD) presents to the subject an item recognition task that requires him/her to respond to aural or visual stimuli that are composed of alphabetic characters or symbols. After a stimulus has been presented, the subject responds by pressing one of two switches to indicate the stimulus is (1) part of a pre-memorized set of letters, or (2) is not part of the set (1). The data collected are the reaction times, in milliseconds, from the onset of a stimulus presentation to the physical response of pressing a button.

When a trial is completed, the subject is given another set of letters to memorize and the sequence is then repeated. Usually four trials are included in a given session with the memory size ranging from one to four letters. Data analysis consists of determining mean reaction time to a number of presentations, and the standard deviation of the response reaction times.

SYSTEM CONFIGURATION

Given the above task, its complexity, number of possible deviations, and timing considerations, a microprocessor was chosen for the main controller. Peripheral devices are manipulated using software located in programmable memory, (RAM). All data collected are temporarily stored in memory and subsequently recorded on a digital cassette tape. When the mission is completed, the tape is retrieved from the real-time controller and taken to a ground-based data analysis computer. Prior to each mission, certain parameters are entered into the ground-based computer and written, along with the operating software, to the cassette tape. These parameters are used by the main controller in order to present variations of the task described.

The system configuration is shown in Figure A-1 and consists of a real-time controller and data collection device, software, and a ground-based data analysis package.

In order to provide for portability, a chassis was constructed to fit into a specific location in the nose of the NT-33A aircraft. The chassis is small enough to fit into a standard off-the-shelf enclosure and light enough to be hand portable. The control panel of the WAD has three connectors that provide all interface signals required for operation. In the portable mode, these connectors provide I/O lines that can be interfaced to various display devices and response keys such as used in many simulators and laboratory environments.

HARDWARE

The Workload Assessment Device (WAD) was designed around the IEEE 696.1/02 buss, (S-100). This buss configuration was chosen for the size of the printed circuit board, availability, and cost. Most S-100 devices manufactured today are reliable and well constructed, and there exists a large base of different peripherals to choose from. Since this system is experimental and cost a major concern, it was not required to meet government/mil specifications for reliability, temperature, and vibration.

The WAD mainframe contains 5 slots that are used for the various peripheral interfaces, memory, and CPU. A single board computer manufactured by Cromemco, Inc., is used for the main controller. It contains a serial I/O port, several parallel I/O lines, real time clock, Read Only Memory (ROM), Programmable Memory (RAM), and all necessary system timing signals. A 16K RAM board is used for program and data storage. The next buss location contains the digital cassette interface which is connected to the NFE Corporation digital tape recorder located on the front panel of the WAD. The fourth slot contains a 16 channel analog to digital (A/D) converter and interface. This unit is connected through cables to the aircraft's analog computer and can be used to monitor up to 6 different control surfaces that will be used in a derivation of the described item recognition task. The fifth position contains the speech synthesizer board. This board contains the new National Semiconductor speech processor IC along with its ROM vocabulary ICs.

SYSTEM OPERATION

When power is applied to the system, a boot program located in ROM loads a file from the cassette tape recorder. This file contains a program that controls all operations of the item recognition task. After loading, the program gains control of the system and waits for a command from the serial port. The experimenter has several options at this time. Usually, he will enter a command for the system to load a specific parameter file from the tape. This file contains all the parameters used in this presentation of the task, such as Inter-stimulus Interval (ISI), Memory Set Size (MSET), use of visual mode versus auditory, etc. After the experimenter enters his selection the subject is presented the task. During the task presentation, all the error scores, reaction times, and other useful data are stored in files of the cassette tape containing all the collected reaction times, error scores, and various other parameters. The program then recycles to the experimenter's console and waits for another command. When the experimental session is over, the cassette tape is retrieved for preliminary data reduction and display.

In order to create a cassette tape containing the operating program and stimulus parameters, a microcomputer is provided that contains 2 floppy disk drives, mainframe, CRT terminal, printer, digital tape recorder, and Disk Operating System (DOS). Located on disks are several user programs that allow the experimenter/scientist to create parameter files. Also located on the disk is a program that contains the operating software for the WAD along with a linker. When the experimenter wants to create a parameter tape he links together the previously created parameter files to be used with the task and the WAD operating software. This newly created link file is then written out to the cassette tape.

During data reduction or analysis, the cassette tape containing the newly recorded data is placed in the tape transport and a loading program is run. This program creates files on the disk containing all the experimental data collected. The experimenter is then able to display the data on the CRT, print it out on the line printer, or submit the data to several data reduction or analysis programs.

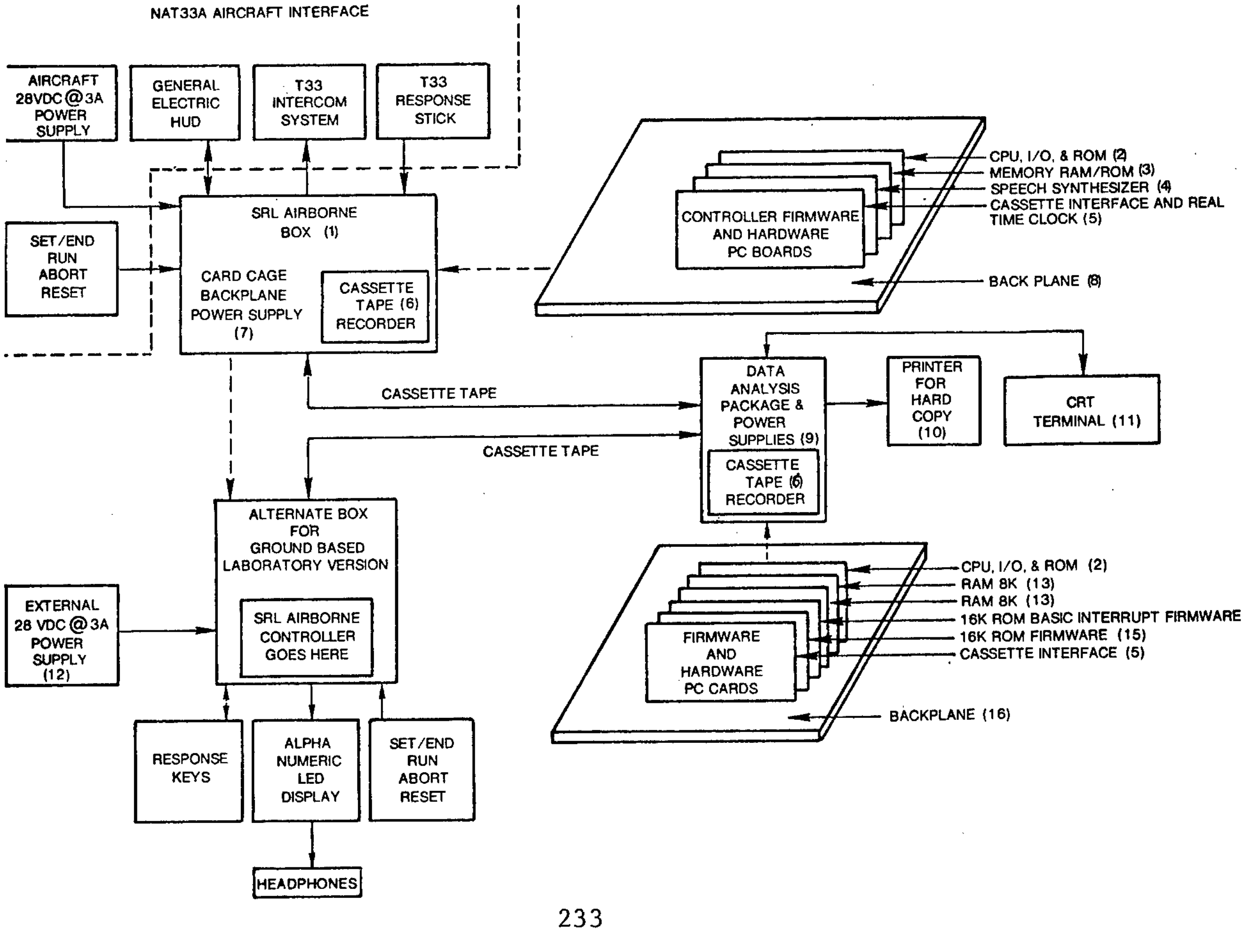
SOFTWARE

The WAD software consists of several programs mentioned above. Most of the hardware drivers and controlling software are written in assembly language, but some of the complex data handling routines are written using Pascal. All of the data analysis software runs under the CP/M disk operating system (DOS). This DOS was chosen because many applications software packages and high level languages are designed to use CP/M for their I/O and file structures.

EXPANDABILITY

Since the main controlling software for the WAD is located at the beginning of each parameter tape and the source is on the floppy disks, it is very easy to modify. The experimenter simply makes his changes using the text editor and recompiles the program. When he makes a parameter tape the new controlling software will be included provided the file name was not changed. This provides the experimenter/scientist with a very versatile system that can be modified for custom applications and has the capability of adding new tasks. Since the peripherals provided are under software control, any sequence of operation can be programmed, thus allowing many different tasks to be included in the data base.

In addition, 16 channels of analog to digital (A/D) converters with interface are being installed. This will enhance the system by allowing it to sample up to six primary control surfaces from the NT-33A, or any other system/simulator. By arranging the data under software control, many derivations of the item-recognition task can be constructed.



Quantification of Pilot Workload via Instrument Scan

J.R. Tole(1,2), A.T. Stephens(1,3), R.L. Harris,Sr.(4), and A. Ephrath(5)
1.Harvard/MIT Division of Health Sciences and Technology
2.Current Address: Biomedical Engineering, Worcester Polytechnic Institute,
3.Current Address: Boeing Co., Renton, Wash.
4.NASA Langley Research Center
5.Bell Labs, Piscataway and M.I.T.

Abstract

This paper describes work in progress on the use of visual scanning behavior as an indicator of pilot workload. The study is investigating the relationship between level of performance on a constant piloting task under simulated IFR conditions, the skill of the pilot, the level of mental workload induced by an additional verbal task imposed on the basic control task, and visual scanning behavior.

The results indicate an increase in fixation dwell times, especially on the primary instrument with increased mental loading. Skilled subjects "stared" less under increased loading than did novice pilots. Sequences of instrument fixations were also examined. The percentage occurrence of the subject's most used sequences decreased with increased task difficulty for novice subjects but not for highly skilled subjects.

Entropy rate (bits/sec) of the sequence of fixations was also used to quantify the scan pattern. It consistently decreased for most subjects as the four loading levels used increased. An exponential equation in task difficulty was found to be a good predictor of entropy rate. When solved for task difficulty, the equation provided an estimate of the level of task difficulty perceived by a subject.

Piloting and number task performance measures were recorded and a combined performance measure was computed. Skill was estimated independently via a method based on pilot experience. These measures were combined with entropy rate to develop a model relating performance, skill, and mental workload. The exponential model fit the data well enough to suggest that this approach has promise in the evaluation of interactions among these variables.

Introduction

The quantification of mental workload in aircraft pilots has been of considerable interest for some time. Perhaps the chief reason for measuring workload is to predict conditions under which task performance will decrement. If such conditions could be accurately predicted, then the nature and temporal sequence of flight procedures and of pilot/aircraft interfaces might be arranged so as to minimize the chances of overload. Quantitative analyses of workload remain elusive however. What one would like is a clear cause and effect relationship between an independent variation in imposed workload and some reliable dependent measure.

The task of flying an aircraft is complex however, and it has been difficult to clarify the functional relationships between various parameters in piloting tasks. The skill a particular individual brings to the piloting task and the nature of the task which is performed can both be expected to influence the "difficulty" of the task. These factors may be further complicated by a shift in the pilot's priorities; (Some tasks may be ignored while others receive full attention).

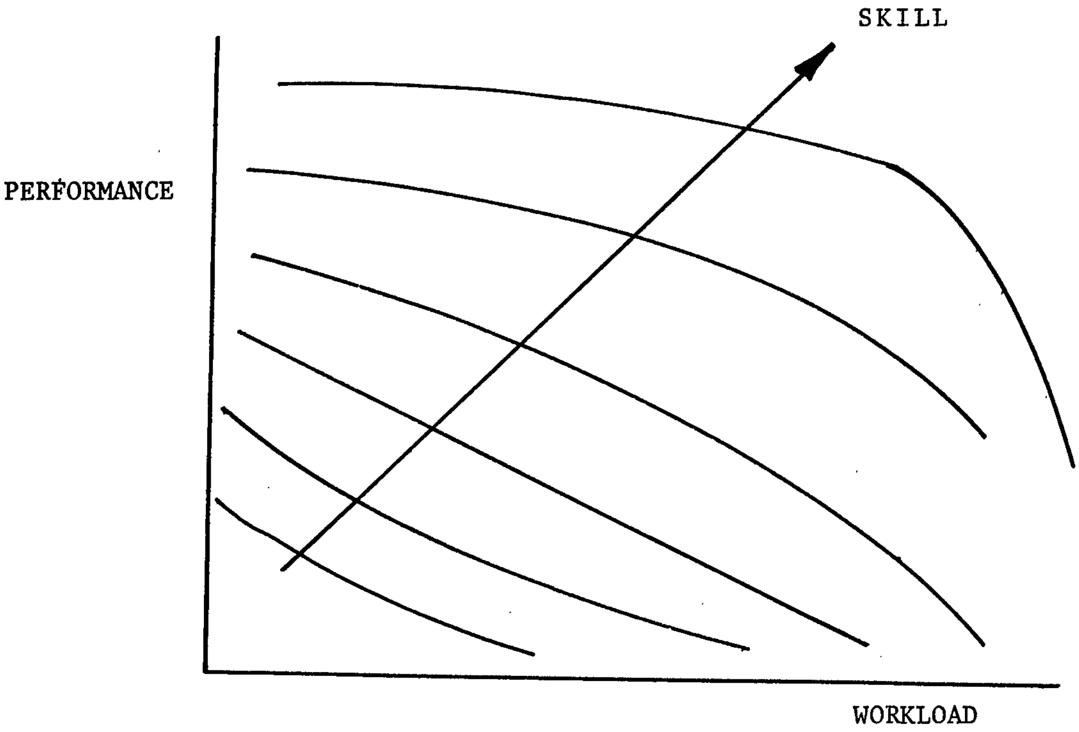


Figure 1. INTUITIVE RELATIONSHIPS BETWEEN PERFORMANCE, SKILL, & WORKLOAD

The problems which such inter-relationships introduce is well illustrated when one attempts to employ task performance as an indicator of workload. All pilots, regardless of skill, can be expected to exhibit poor performance if the loading level is excessive. The overload situation is relatively easy to assess, however, using subjective techniques. Situations which involve intermediate to high levels of loading would seem to be the ones of more practical concern; i.e., one is concerned with minimizing the chance of a high workload approaching an overload situation. Intuition suggests that the level of skill of the pilot may influence the performance vs workload relationship for intermediate or marginal loading levels. A pilot of high skill would be expected to maintain "better" performance than a novice flyer under any loading condition short of overload. This intuitive concept is illustrated graphically in figure 1.

The research described here uses this graphical representation of the performance/skill/workload relationships in order to pose a number of testable hypotheses. It will be suggested shortly that instrument scan may be an indicator of workload and/or skill in certain types of flight situations, a suggestion supported by both qualitative and quantitative results. In addition, if a measure of workload based on instrumnt scan is combined with independent measures of pilot skill and performance, then a model of the hypothetical relationships in figure 1 may be developed and tested.

Visual Scanning Behavior

The pilot has many sources of information input but the most important one during instrument flight is probably the visual pathway. Under instrument flight conditions, some sensory inputs may even provide false information such as vertigo which results from conflicting visual and vestibular information. The pilot obtains information concerning aircraft state by cross-checking or scanning the flight instruments. The exact method of scanning the instrument panel varies from pilot to pilot but there are some basic features common to a "good" scan pattern. Indeed, it was the early study by Fitts and his associates on instrument transitions which led to the familiar "T" arrangement of the major flight instruments (Jones, et.al., 1946).

A fundamental notion n the present work is that a repetitive piloting task will invoke a regular visual scan (spatial/temporal pattern of eye movements) during instrument flight. If this notion is correct, then it may be postulated that external factors such as noise, interruptions, and fatigue which interfere with the piloting task may produce measurable changes in the scanning behavior. Such a measure would be particularly attractive for quantifying workload since it would be both non-invasive and objective.

Experimental Design

A series of experiments is being carried in order to carefully examine these ideas. The basic experiment is described in detail elsewhere (Tole, et al. 1982) and only the salient points are repeated here. The experiments described were performed at the NASA/Langley Research Center, Flight Management Branch, in Hampton, Virginia, making use of their flight simulator and oculometer facilities (Middleton, et.al., 1977).

Three factors were manipulated in the experiments: 1) a piloting task requiring a stereotyped scan path, 2) a verbally presented mental loading task, and 3) a workload calibration side task.

We sought a representative constant piloting maneuver which might be realistically expected to occur for periods of up to 10 minutes in actual flight. This run length was chosen as an estimate of the minimum amount of time required to provide a sufficient number of instrument fixations to satisfy the assumption of steady state conditions. The Instrument Landing System (ILS) approach is often chosen as the piloting task in studies of workload (Waller, 1976; Krebs and Wingert, 1976; Spady, 1977). Hovever, the ILS approach represents a constantly changing task difficulty as touchdown is approached (especially due to increases in Glide slope sensitivity and cost of error for course deviation). This variation in the primary task loading makes it difficult to accurately control the amount of mental workload on the pilot as an independent variable. It was decided that a scenario in which glide slope sensitivity and heading were held constant would allow the piloting task difficulty to remain relatively constant for a long period, but nevertheless be more or less realistic.

A desktop general aviation instrument flight simulator (Analog Training Computers ATC-510) was used to simulate these flight manuevers. The ATC-510 is a procedures trainer for light, single engine, fixed pitch prop, fixed gear, IFR equipped aircraft. The simulator was equipped with a turbulence level control which was set to the first level above calm conditions in order to force some pilot vigilance on the flight task.

Pilot lookpoint on seven instruments (Attitude Indicator 'ATT', Directional Gyro 'DG', Altimeter 'ALT', Vertical Speed Indicator 'VSI', Airspeed 'AS', Turn and Bank '*B', and Glide Slope/Localizer 'GSL') was measured using a Honeywell oculometer system which has been substantially modified by NASA Langley Research Center (Middleton, et.al., 1977). This device is non-invasive and allows the user to determine the time course of eye fixations on instruments employed by the pilot and the dwell time of each fixation to the nearest 1/30 sec.

The mental loading task was chosen so as not to directly interfere with the visual scanning of the pilot (i.e. the task would not require the pilot to look away from the instruments) while providing constant loading during the maneuver. The task used required the pilots to respond to a series of evenly spaced three-number sequences (Wittenborn, 1943) presented to them audibly by means of a speaker. The pilot was told that he must respond to each three-number sequence by indicating either "plus" or "minus" according to the algorithm : first number largest, second number smallest = "plus" (e.g. 5-2-4), last number largest, first number smallest = "plus" (e.g. 1-2-3), otherwise, "minus" (e.g. 9-5-1).

The mental workload experienced by the pilot is inversely proportional to the intervals between number sequences. This relationship is given by the following equation which is arbitrarily chosen:

(1) TD = 1/interval between # task

where TD is equal to imposed task difficulty. The four loading levels used in the current experiments were intervals of continuous silence (i.e. no-numbers presented), ten, five, and two seconds which have corresponding task difficulties of 0.0, 0.1, 0.2, and 0.5, respectively.

Numbers were generated by a computer controlled speech synthesizer. This allowed automated scoring of task accuracy, calculation of response reaction times, and the possibility of temporal correlations of visual or other responses with the verbal stimulus. The probabilities of occurence of "+" and "-" sequences were each 0.5. The pilot was instructed to give the number task priority equal to that of the piloting task as if the verbal questions represented a constant rate of radio communication. Performance was recorded by having the pilot press a 3-position rocker switch mounted on the yoke up for plus and down for minus.

The amount of mental loading imposed on the pilot by the number task was calibrated using a side task (Ephrath, 1975). The runs made with the side task were not used in the scanning analysis, however, due to the alteration of normal scanning caused by the task. The results (Tole, et.al., 1982) from these runs confirmed the relative difficulty of the various number intervals.

A microprocessor development system (Burns, et.al, 1980) was used for both stimulus presentation and data collection and analyses.

Performance Measures

Several variables were obtained from each of the twotasks in order to allow the computation of performance scores. The scores developed ran between 0 percent and 100 percent with 100 percent being obtained if the pilot never deviated from the intended path in space on the piloting task, and if all number task sequences were answered correctly for the mental loading number task. The scores from the piloting and the mental loading tasks were then combined to provide a performance measure to be used in the validation of proposed performance/skill/workload model.

The scoring measure for the number task was computed as given below.

$$(2) \qquad \#TP = \frac{(TOT - WRO - MIS)}{TOT} \times 100\%$$

where

TP = mental loading number task performance
TOT = total number of stimuli presented
WRO = number of incorrect responses
MIS = number of missed responses

This score was 100 percent if the pilot answered every sequence correctly and zero percent if a pilot either answred incorrectly or missed all of the stimuli presented. Most subjects score nearly 100% on this task if they have nothing else to do simultaneously.

The raw data available for scoring performance on the piloting task were the errors from the intended track for the glide slope and localizer courses. Discussions with several highly skilled pilots revealed that accuracy of tracking the glide slope and localizer might not provide a complete performance picture. These pilots were willing to trade off "smoothness" when the loading task became more difficult; i.e. the pilot may perform the piloting task to the same level of accuracy, as far as deviations from a designated path are concerned, on two different runs but produce two very different ride qualities for these runs. One possible measure for smoothness could be the frequency of oscillation around the intended path. The higher this frequency is, the less "smooth" the ride becomes. It was arbitrarily assumed that a smooth ride would contain frequecies mostly less than 0.1 Hz. Under this assumption, measurement of the spectral component of the aircraft dynamics above 0.1 Hz. would indicate any decrement in the ride quality.

In order to examine this measure, the power-spectral density (PSD) of the course deviations was computed. The bandwidth of the calculated PSD was 2.5 Hz. The "power" within a band of frequencies may be determined by integrating the PSD over that band (Schwartz, 1959). We chose to consider the % of the spectral power which was located in the band from 0.1 to 2.5 Hz. This was calculated by subtracting the power contained in the band from 0 to 0.1 Hz (assuming that the D.C. component was first removed) from the total power in the spectrum and multiplying by 100%. This % of the PSD was computed for both the glide slope and the localizer and combined wth the two RMS measures to provide four candidate variables to be included in a performance score for the piloting task.

Since the pilots were instructed to give equal priority to the piloting task and the mental loading number task, both were included in the development of a combined performance score. While a weighting of 0.5 might have been assigned to each task, it was decided to leave the weighting free to allow the model fitting procedure to determine the relative weights. A linear relationship between all of the terms was assumed and the form of the equation became:

$$P = CONST + a(\#TP) + b(RMS/GS) + c(RMS/LOC) + d(\%PWR/GS) + e(\%PWR/LOC) \quad (3)$$

where
P = combined performance measure
CONST = constant term
TP = mental loading number task performance
RMS/GS = RMS error from glide slope track
RMS/LOC = RMS error from localizer track
%PWR/GS = percent of power from the power-spectral density for the glide slope greater tan 0.1 Hertz
%PWR/LOC = percent of power from the power-spectral density for the localizer greater than 0.1 Hertz

Estimation of Pilot Skill levels

In order to assess the effects of skill on performance and mental workload, an independent quantitative measure of skill was needed. A model of pilot skill based on experience factors was used for this purpose (Hollister, et al, 1973). This model was developed in order to predict the current level of skill of pilots flying light, single engine aircraft.

$$\text{Skill} = 1.42 + 0.25(\text{recency}) + 0.73(\log(\text{total time})) - 0.030(\text{years certified}) + 0.15(\log(\text{time in type})) - 0.0088(\text{age}) + e \quad (4)$$

where
Skill = score reflecting relative piloting performance
recency = number of flight hours in past 30 days
total time = total number of flight hours
time in type = total number of hours in light single engine aircraft
years certified = time in years since last certificate orating
age = subjects's age in years
e = residual variance not explained by the model

A raw skill score was calculated for each of the pilot subjects using the model. The pilot with the highest resulting skill score was then used to normalize all of the scores so that skill levels would range between 0% and 100%. Eleven subjects ranging in skill from NASA test pilots to non-pilots participated in the experiments. The relative skill scores for the subjects are given in Table I.

NASA PILOT#	SKILL SCORE
3	100%
4	85
11	77
13	53
15	39
6	37
12	33
14	32
8	22
7	15
16	13

TABLE I.
Relative Skill Scores of Subjects based on Equation 4

Though care must be taken when applying an equation such as this in a different set of experimental conditions, the overall rank ordering of the pilots by this method is probably accurate as it generally agreed with subjective rating of the pilot's skills by experienced observers at the NASA/Langley Research Center.

Conduct of the Experiments

Each session consisted of four 10-minute runs with a 5-minute break between each run. The difficulty of the mental loading task would start at no numbers for the first run and increase to 2-sec intervals by the fourth run. Some subjects participated in two sessions, one without and one with the side task. Each subject was allowed to practice all three tasks until he felt comfortable with them.

Preliminary Results

Instrument dwell time histograms and the frequency of usage of different sequences of instrument fixations were both affected by the loading task. Both results are reported in detail elsewhere (Tole, et.al., 1982) and , only the major points are mentioned here. An increase in dwell time with increase in mental loading was observed in all subjects. This is illustrated in figure 2. Novice subjects generally had much longer dwell times under increased load than did skilled pilots. (Relative skill levels are given in Table I above.) The fixation sequences of the pilot's instrument sans were analyzed, and the percentage occurrence of the ten most frequently occurring sequences were also analyzed. These results

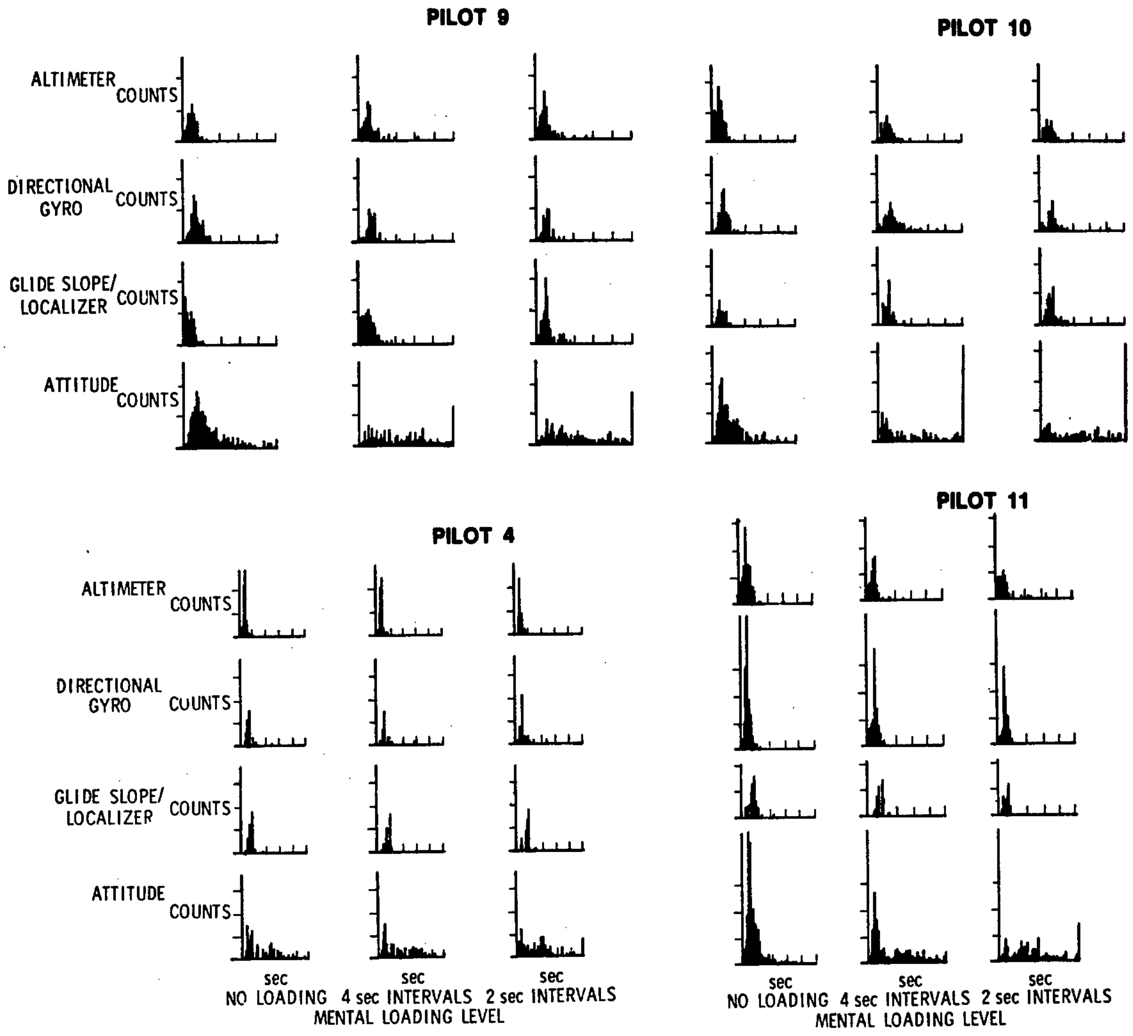


Figure 2. DWELL TIME HISTOGRAMS FOR TWO SKILLED PILOTS (#4 & #11) AND TWO NOVICE PILOTS (#9 & #10) UNDER VARIOUS. LOADING CONDITIONS

indicate that: 1) skilled pilots use a higher percentage of their ten most frequently occurring sequences than do novice pilots and 2) the scan pattern of the novice subjects were affected more by the increase in mental loading than were the patterns of the highly skilled pilots. This result is shown in figure 3.

A more general method of quantifying the scan

Traditionally, much of the quantitative analysis of scanning patterns has employed Markov transition probability matrices (Stark and Ellis, 1981; Krebs and Wingert, 1976). Such matrices do describe the predominant patterns in the scan via the relative sizes of transition probabilities but it is either extremely unwieldy or impossible to compare two of these matrices for different experimental conditions. One of the major goals of this research is the identification of general methods for the study of scanning behavior. To be most useful the method should be independent of the number and arrangment of instruments. The nature of eye-point-of-regard data (sequential instrument and dwell times) obtained from the oculometer suggests several methods from information theory which may have this generality.

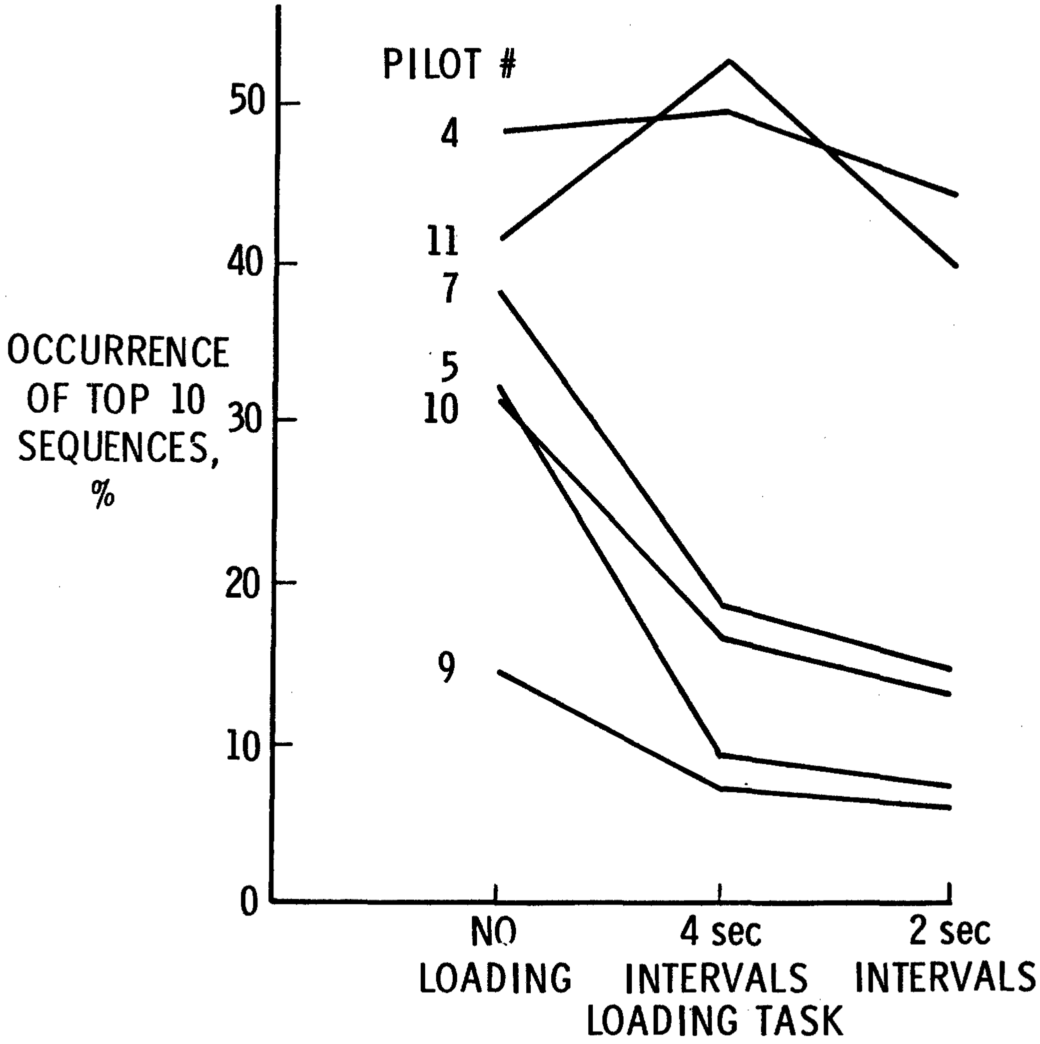


Figure 3. PERCENT USAGE OF LENGTH 4 SEQUENCES UNDER VARYING LOAD (TYPICAL SEQ : ATT - DG - ATT - ALT)

The piloting task in the current experiment is such that the pilot's scan can only lie on one of the 7 specified instruments although each fixation may be of arbitrary duration. The time history of fixations has a form which is similar to that of a communications system which can assume 7 discrete states with a varying duration in each state. The orderliness of such a system is related to the probabilities with which it occupys its different states. A system which always occupied the same state or always made the same transitions between states would thus be quite orderly. In the case of instrument scan, these situations would be paralleled by staring and by a stereotyped scanpath respectively.

This concept of system order may be stated compactly using the mathematical form for entropy from information theory. The entropy of a sequence is defined as (Shannon and Weaver, 1949):

(5) $$H_o = -\sum_{i=1}^{D} p_i \log_2 p_i$$

where

H_o = observed average entropy

p_i = probability of sequence i occurring

D = #of Different sequences in the scan

In the case of the instrument scan, entropy has the units of bits/sequence and provides a measure of the randomness (or orderliness) of the scanpath. The higher the entropy, the more disorder is present in the scan. The maximum possible entropy is constrained by the experimental conditions (see below). The entropy measure uses the same probabilities which are present in transition matrices, but it yields a single, more compact expression for the overall behavior of the probabilities rather than presenting them each individually. This method appears to afford some generality and has been the focus of our recent efforts.

To implement this method, each of the instruments to be examined was given a number. Then a sequence of these numbers was stored as the pilot scanned the instrument panel together with the dwell time for each fixation. While sequences of up to length 4 were considered in preliminary analyses, the most detailed study was made on sequences of length 2. The remainder of the discussion here applies to the results for length 2 sequences. Details of themethodolgy are given elsewhere (Stephens, 1981).

It can be shown that the observed entropy for the instrument scan is related to the total number of fixation sequences (L, defined with equation 7 below) observed during a run. In order to compare entropies from the scans of different pilots for different run lengths, each estimate of entropy had to be corrected for L and normalized to its maximum possible value, Hmax. Hmax may be calculated as follows. In the most general case, M instruments may be arranged in some arbitrary fashion on the cockpit panel. For a given number of instruments, M, and sequence length N, the maximum number of different fixation sequences is given by:

(6) $$Q = M(M-1)^{N-1}$$ = maximum number of sequences of length N

The number of bits required to uniquely encode all Q possible sequences is log2 Q. The magnitude of this latter number also represents Hmax of the visual scan for the number of instruments an sequence length being considered. For example, with 7 instruments the value of Q for sequences of 2 instruments is 56 which yields a corresponding Hmax = 5.8.

The normalized value of H may then be calculated from:

$$(7)\quad Hcorr = Ho * \frac{Hmax}{\log_2 L}$$

where

$L = R-N+1$ = number of sequences in a run
R = number of fixations in a run
N = sequence length (N = 1,2,3, or 4)

While entropy should help to explain the orderliness (or lack thereof) of the scanning pattern, the development presented up to this point does not include the fact that the dwell time for each fixation is different. From the preliminary results on instrument dwells, it appears rather clear that dwell times can be markedly affected during high mental loading. In order to include the effect of time in our measure, a term for entropy rate was defined as:

$$(8)\quad Hrate = Ho/t$$

where Ho is the entropy for the system given by 7 and t = smallest interval in which a transition may occur.

In practice, the calculation of Hrate was an average value given by the following:

$$(9)\quad Hrate_{avg} = \sum_{i=1}^{D} Hcorr_i / DT_i$$

where

$Hcorr_i$ = Normalized entropy for ith sequence
DT_i = Average Dwell time for ith sequence
D = # of different fixation sequences

It is helpful to estimate the maximum value which Hrate might assume. This may be calculated using the maximum for entropy determined above together with dwell time statistics for the various instrument sequences in the scan. While it is possible for pilots to make rather rapid glances (with dwell times of 100 msec or less) at their instruments (Harris and Christhilf, 1980) a fixation rate this high (10 fixations/sec) rapidly leads to oculomotor fatigue. A morerealistic average value is probably about 2 fixations/sec or less for a long period of instrument scan (say > 10 sec).

Using 0.5 sec/look (2 fixations/sec) as the average dwell interval, the maximum entropy rate for sequences of length 2 is calculated to be

$$Hrate_{max} = 5.8/0.5 * 2 \text{ fixations/seq.} = 6 \text{ bits/sec}$$

This number represents an upper bound. Since we suspect that the pilot must have some regularity in his or her scan, the numbers we would expect to obtain under actual flight conditions will probably be lower. The observed average Hrate for the current experiments was on the order of 1 bit/sec. A tendency to stare under increased load should be reflected by decreased entropy and increased fixation times making Hrate tend toward lower values under such conditions. Figure 4 plots Hrate vs number Task Difficulty for all pilots except 12 and 8.

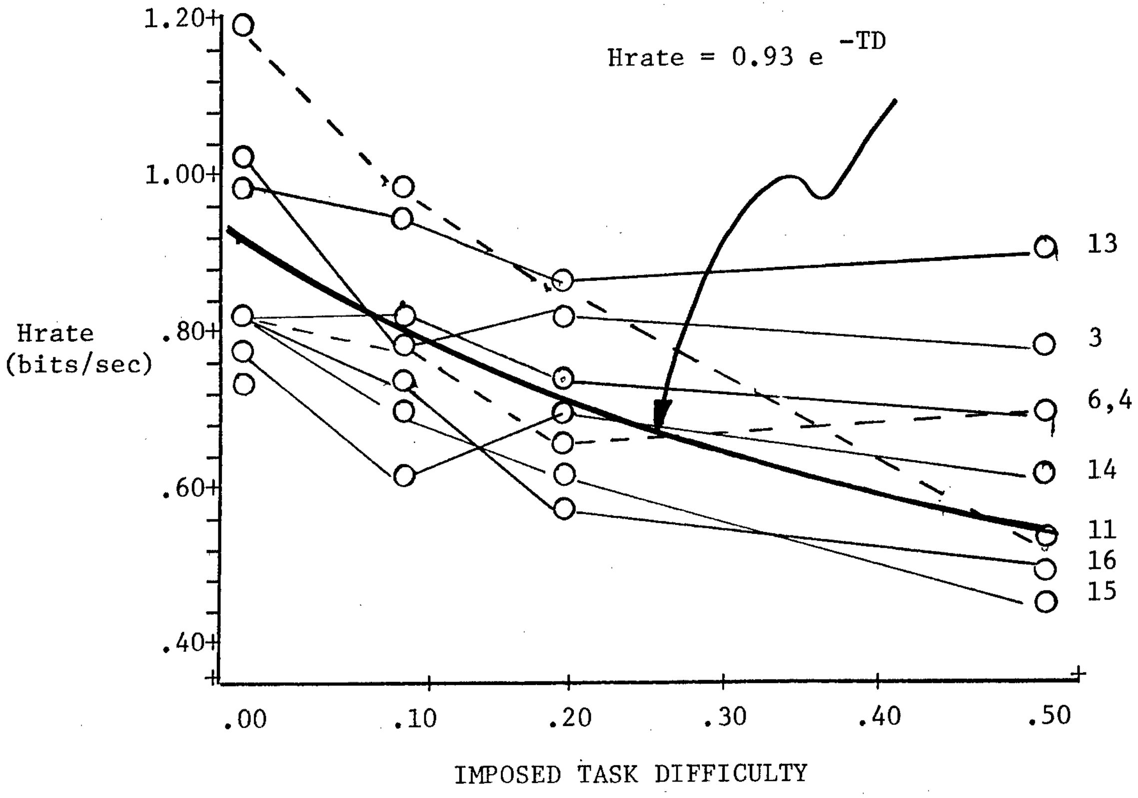


Figure 4. ENTROPY RATE ON LENGTH 2 SEQUENCES vs. IMPOSED TASK DIFFICULTY

A trend toward lower entropy rate with higher task difficulty may be seen. A two-way analysis of variance was performed for the entropy rate data from nine pilots on levels of task difficulty and between subjects. F tests allowed rejection of two null hypotheses: equality of mean Hrate at all loading levels ($p < 0.01$) and equality of mean Hrate between subjects ($p < 0.01$). All six combinations of level differences in mean Hrate were found to be statistically significant (T-test $p < 0.05$). Thus Hrate was chosen to map from scanning behavior into task difficulty (i.e. workload).

The model used expresses Hrate as an exponential function of TD.

(10) Hrate = 0.9279 EXP(-TD)

This equation was obtained via a regression analysis based on the data from seven of the pilots with a coefficient of determination, R-squared, = 97.3%. This equation may be solved for task difficulty withthe following results:

$$(11)\quad TD = -(0.06 + \ln Hrate).$$

This expression can then be used to predict the level of TD for a new subject under the conditions of the experiment reported here.

Model Development and Verification

One of the major goals of this work was the development of a model relating performance, skill, and mental workload. The ultimate goal is the prediction of performance given estimates for skill and scanning parameters. A model relating performance, skill, and mental workload may be postulated from the empirical relationship shown in figure 1. Construction of the model should, in fact, aid in determing whether such empirical expressions are valid. The model chosen was an exponential form:

$$(12)\quad P = P(0) - EXP((TD/Skill)^{2})$$

This equation may be rearranged as follows:

$$(13)\quad EXP((TD/Skill)^{2}) = P(0) - P$$

which states that the exponential term is equal to the difference in te performance at the no-loading level P(0) and the performance at the present level of mental loading P. Using the values for the level of skill and task difficulty calculated in equations 4 and 11 respectively, the left hand side of the equation may be computed. The right hand side of the equation must be expressed in terms of measurable performance indicators.

Expanding the right side of (13) yields

$$\begin{aligned}(14)\quad P(0) - P = {} & a(\#TP(0) - \#TP) + b(RMS/GS(0) - RMS/GS) \\ & + c(RMS/LOC(0) - RMS/LOC) + d(\%PWR/GS(0) - \%PWR/GS) \\ & + e(\%PWR/LOC(0) - \%PWR/LOC)\end{aligned}$$

A multiple regression analysis was then performed on equation 13 using values for each of these measures recorded during the experiments.

The data from seven pilots was used for model development, while that from three other subjects was used for model verification. One pilot's performance data was discarded due to equipment malfunction.

The results of the first attempt at regressin indicated that the coefficient of the %PWR/LOC term could not be differentiated from zero based on a Student's T-test. This variable was eliminated from equation 13 and the analysis was repeated. This regression yielded non-zero values for the coefficients a through d, and included a constant term. The resulting equation was:

$$\begin{aligned}(15)\quad EXP((TD/Skill)^{2}) = {} & 1.4483 + 0.0351\#TP(0) - \#TP) \\ & + 0.1765(RMS/GS(0) - RMS/GS) - 0.0366(RMS/LOC(0) - RMS/LOC) \\ & + 0.0377(\%PWR/GS(0) - \%PWR/GS)\end{aligned}$$

This analysis had an R squared value of 76.6 percent and an F-ratio of 12.28 ($p < 0.01$). The coefficients determined for 15 may now be used in equation 3 which becomes

$$\begin{aligned}(16)\quad P = {} & 1.4483 + 0.0351\#TP) + 0.1765(RMS/GS) \\ & - 0.0366(RMS/LOC) + 0.0377(\%PWR/GS).\end{aligned}$$

These coefficients provide the relative weightings for each of the performance terms but they need to be scaled in order to provide the proper characteristics for the equation. If each of the terms were at their maximum value, that is 100 percent, then the combined performance measure should also equal 100 percent. However, using the coefficient this 100 percent, each coefficient must be multiplied by 100./22.72 = 4.40. The modified performance equation becomes:

$$\begin{aligned}(17)\quad P = {} & 6.3750 + 0.1545\#TP) + 0.7769(RMS/GS) - 0.1611(RMS/LOC) \\ & + 0.1659(\%PWR/GS)\end{aligned}$$

A plot of this fuction versus the task difficulty, obtained from equation 11, is provided in Figure 5.
It was hoped that these curves would resemble those given in the hypothetical plot in Figure 1 and for some of the pilots, a general overall downward trend is present. Even though the curves do not match the hypothetical ones exactly, there are some common features between them. First of all, the curve for the lowest skilled pilot 7 is seen to decrease much more rapidly than the curves forthe more highly skilled pilots (3, 11; the two points for 3 are for the third and highest levels of mental loading respectively).

To test this model's value as a predictive tool, the data from three subjects not included in the model determination, were substituted into equation 17 and plotted versus perceived task difficulty in Figure 6. Pilots 12, 8, and 16 produce some interesting, if not consistent results. The three points of pilot 12, and pilot 16 are for the second, third, and highest loading levels. All three pilots show a net decrease in performance between their lowest and highest task difficulties even though they accomplished this decrease in very different ways. Pilot 8 appears to be the closest to the theoretical model with his sharp decrease in performance over a very small task difficulty increase. Pilot 16, on the other hand, appears to be decreasing at an exponentially decreasing rate as opposed to the model which predicts reasing performance at an exponentially increasing rate. Pilot 12 increases performance sharply between his second and third runs and then decreases just as sharply between the third and fourth runs.

Since the choice of the exponential model for performance/skill/workload was arbitrary, two other forms for the model were also examined. These were circular and linear models and neither was as good at fitting the data as the exponential and hence were abandoned.

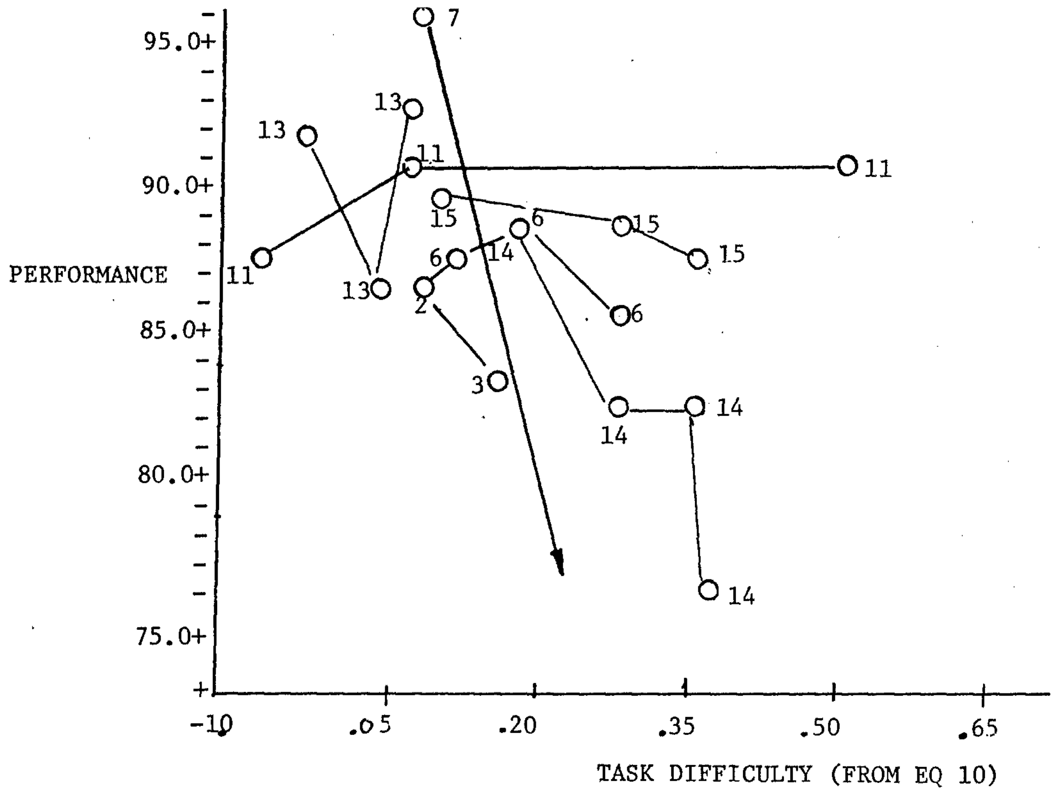


Figure 5. Combined performance (from model) perceived task difficulty for 7 pilots used in model development

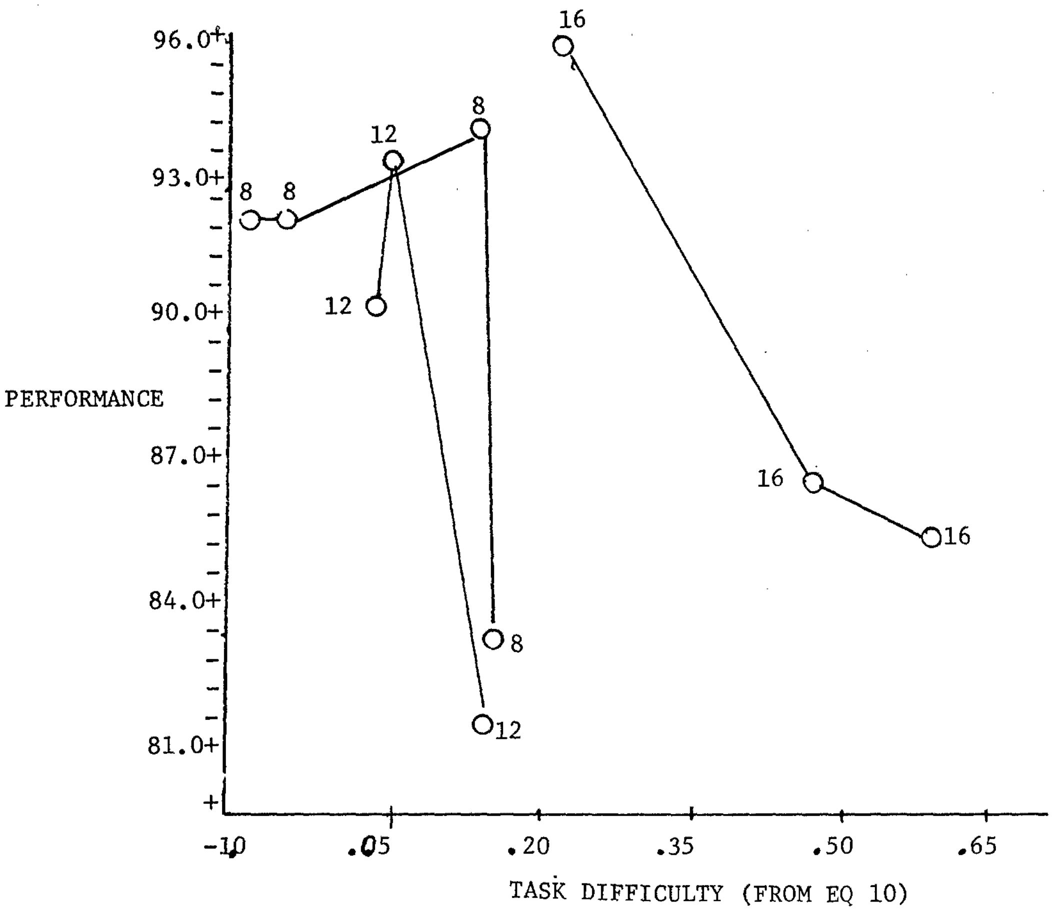


Figure 6. Combined performance vs. task difficulty for 3 test cases of model

The models described here are still under development and work is in progress to repeat the experiments described here and to apply this methodology to other instrument flight scenarios.

Summary

This paper presents some of the findings frm a set of experiments designed to explore the relationship between performance, skill, and visual scanning behavior of aircraft pilots under varying levels of mental workload. Instrument fixations were recorded as a group of pilots with widely varying levels of skill simultaneously performed a constant instrument flight task and a verbally presented loading task with 4 discrete levels. Initial results indicate a tendency of lesser skilled pilots to stare at the primary instrument as loading is increased and to alter the frequency of usage of different scan paths. Skilled pilots demonstrated much less change on both of these measures.

A major finding of the research suggests that under relatively constant instrument flight conditions the entropy rate of the visual scan path may be a useful measure of the level of mental workload induced by a constant rate verbal task. This measure of workload was combined with independent estimates of performance on the piloting and verbal tasks and of pilot skill. An exponential model relating these factors was developed and has undergone preliminary tests. The model helps provide insight on the intimate connections between a particular workload measure and operator skill and performance strategy.

Acknowedgements: This work was supported by NASA Co-operative Agreements NCC 1-23 and NCC 1-56. The verbal loading task was suggested by N. Moray. The use of entropy as a measure of the visual scan was suggested by A. Natapoff. The technical assistance of M. Goode is gratefully acknowledged.

References

Atteneave, F., Application of Information Theory in Pyschology, Holt, Rinehart, and Winston, 1959.

Burns,S.K., Dowling,N., Peterson,S., Pettyjohn,R., Schluter,P., Tole,J., Valvano,J., and Mark,R. "Experience With a Modular Microprocessor Base for Medical Instruments", IEEE-EMBS Meeting, Denver, October, 1980.

DeMaio, J., Parkinson, S., Leshotz, B., Crosby, J., and Thorpe, J., "Visual Scanning: Comparisons between Student and Instructor Pilots", Air Force Human Factors Resources Laboratory Report AFHRL-TR-76-10.

Dick, A.O. "Instrument Scanning and Controlling: Using Eye Movement Data to Understand Pilot Behavior and Strategies", NASA CR-3306, September, 1980.

Ellis, S.R. and Stark, L. "Eye Movements during the Viewing of Necker Cubes", Perception, 7:575-581, 1978.

Ephrath, A.E., "Pilot Performance in Zero-Visibility Precision Approach", Ph.D. Thesis, Dept. of Aero. and Astro., M.I.T., June, 1975.

Harris, R.L., Sr., and Christhilf, D., "What do Pilots See in Displays?", presented at the Human Factors Society Meeting, Los Angeles, October, 1980.

Harris, R.L., Sr. and Mixon, R., "Advanced Transport Operation Effects on Pilot Scan Patterns", proceeding of the Human Factors Society, 23rd annual meeting, Boston, Massachusetts, 1979, pp 347-351.

Holland, M.K. and Tarlow, G. Blinking and mental load. Psychological Reports, 1972, 31, 119-127.

Hollister, W., LaPointe, A., Oman, C. and Tole, J. "Identifying and Determining Skill Degradations of Private and Commercial Pilots", FAA Report FAA-RD-73-91, June, 1973.

Jones, R.E., Milton, J.L., and Fitts, P.M., "Eye Fixations of Aircraft Pilots: A Review of Prior Eye Movement Studies and a Description of a Technique for Recording the Frequency, Duration, and Sequence of Eye Fixations during Instrument Flight", USAF Tech Report 5837 (AT I 65996), 1946.

Kennedy, J.P. Time-sharing effects on pilot tracking performance. Master's thesis, Naval Postgraduate School, Monterey, California, September, 1975. (AD AO16 378).

Krebs, M.J., and Wingert, J.W., "Use of the Oculometer for Pilot Workload Measurements" NASA CR-144951, Feb, 1976.

Middleton, D.B., Hurt, G.J. Jr., Wise, M.A., and Holt, J.D. "Description and Flight Tests of an Oculometer", NASA Technical Note NASA TN D-8419, June, 1977.

Moray, N., Ed., Mental Workload, Theory and Measurement, Plenum, 1979.

Noton D. and Stark, L., "Eye Movements and Visual Perception", Scientific American, 244(6):35-43, June, 1971.

Schwartz, M., Information Transmission, Modulation, and Noise, McGraw Hill, New York, 1959.

Shannon, C.E. and Weaver, W. THE MATHEMATICAL THEORY OF COMMUNICATION, Univ. of Illinois Press, 1949.

Spady, A.A., "Airline Pilot Scan Patterns During Simulated ILS Approaches", NASA Technical Paper 1250, October 1978.

Stark, L. and Ellis, S., "Scanpaths revisited: Cognitive Models Direct Active Looking", in Fisher, D., Monty, R., and Senders, J., Eye Movements Cognition and Visual Perception, Erlbaum, Hillsdale, N.J., 1981.

Stephens, A.T., Instrument Scan, Performance, and Mental Workload in Aircraft Pilots, S.M. Thesis, Dept. of Aero. and Astro., M.I.T., Sept., 1981.

Tole, J.R., Stephens, A.T., Harris, R.L.,Sr., and Ephrath, A., "Visual Scanning Behavior and Mental Workload in Aircraft Pilots", Aviation Space and Environmental Medicine, Jan., 1982.

Waller, M.C. An investigation of correlation between pilot scanning behavior and workload using stepwise regression analysis. Hampton, Virginia: NASA Langley Research Center, NASA TM X-3344, March, 1976.

Weir, D.H. and Klein, R.H. Measurement and analysis of pilot scanning behavior during simulated instrument approaches. Proceedings of the 6th Annual NASA-University Conference on Manual Controol, Wright-Patterson AFB, April, 1970, 83-108.

Wittenborn, J.R., "Factorial Equations for Tests of Attention", Psychometrika, 8(1):19-35, March, 1943.

Question a.

Considerable research along similar lines was done under ONR/NASA sponsorship several years ago at STI (e.g. see NASA CR 1569 and ONR reports "STI TR 163-1 and 183-). Invokes control theory analysis to show that scan patterns are not completely random, but have predictable (explainable) correlations with controlled element and task demands. (Data show similar effects as yours).

Answer a.

There certainly are some interesting parallels between our work and earlier studies at STI as in NASA CR 1569. Both efforts reveal an observable change in scanning behavior with varying task difficulty. The experimental conditions are somewhat different, however, in that the STI work uses a "critical side tracking track" which requires an alteration in the scan. Our method (verbal task if varying difficulty) does not in itself require an altered scan path for its successful performance. As the critical task difficulty increases the swell times become shorter. For increased verbal loading in our experiments, the dwell times become longer.

While these two findings are not directly comparable, they do point out the potential utility of instrument scan in the measurement of behavior of pilots and the need for great care in the interpretation of scanning data within the context of a particular experiment.

Question b.

(a "nit") Why use the arcane term "entropy" and "ℓ—" rate when the current term (circa 1960's and on) is "transinformation index" and rate (e.g. used by Ames references since 1960's)?

Answer b.

The use of the word "information" would be misleading in the context of our experiments since we do not currently attempt to quantify the amount of information the pilot is obtaining from his displays. Rather, we are concerned for the moment only with the orderliness of the scan pattern. The method used to quantify the order in the scan was the mathematical form of entropy as presented in the original works on information theory (Shannon & Weaver, 1949). Entropy seems a clear enough term; "transinformation" on the other hand suggests a broader meaning than we intend in our work reported here.

HELICOPTER PILOT PERFORMANCE AND WORKLOAD AS A FUNCTION OF NIGHT VISION SYMBOLOGIES

E. James Hartzell

U.S. Army Aeromechanics Laboratory
Moffett Field, CA

Abstract

In the near future military helicopter pilots will be required to fly night time nap-of-the-earth (NOE) missions using helmet-mounted displays of their flight environment. These displays will be composed of an infrared video representation of the outside world and flight control information, in a symbology format, superimposed on the visual scene. The piloting task under such conditions must be classed as extremely demanding in terms of performance requirements, safety, and workload. A study was undertaken to investigate several human factors questions of man-machine integration mediated through such displays and symbologies. A full six-degree-of-freedom motion simulation of an advanced U.S. Army helicopter flying a night NOE scenario was conducted. The dynamic visual scene was obtained from a video picture of a terrain board and three-candidate computer generated flight control symbologies were video-mixed with the scene. Six experienced helicopter pilots were employed as subjects and trained to fly a scenario incorporating multiple precision hover maneuvers which varied in difficulty and task loading. The experiment was designed to assess pilot performance, training requirements, and workload as a function of the three symbologies. Workload is essentialy a cognitive function and is, therefore, difficult to measure objectively though Wierwille and Williges reviewed some 28 different techniques which have been attempted. Of the techniques available, Time Estimation was chosen as the most unobtrusive and least problematic to implement. In a combat environment, helicopter pilots are required to estimate the period of time that they can safely remain exposed or the period before exposure is executed. Time estimation periods were, therefore nested in the current experimental scenario which agreed with and supported external validity. The results of the time estimation technique used in conjunction with this experiment supported the predicted levels of difficulty designed into the flight scenario. Time estimation provided a relative scale of workload between the hover maneuvers and suggests that there is no significant difference in workload between the three symbology types studied.

An Investigation of the Effects of an Isometric Side-Stick Controller on Pilot Workload for Helicopter Terrain Flight (Progress Report)

John C. Hemingway
NASA-Ames Research Center
Moffett Field, California

Edwin W. Aiken and Christopher L. Blanken
Aeromechanics Laboratory
U.S. Army Research and Technology Laboratories (AVRADCOM)
Ames Research Center
Moffett Field, California

Abstract

The design and tentative results of a moving-base simulator experiment conducted to assess the effects on pilot workload of replacing one or more of the standard helicopter controllers with an integrated isometric side-stick controller are described. Using the NASA-Ames Vertical Motion Simulator, various low-altitude helicopter maneuvers typical of daytime terrain flight are simulated to investigate the effects of parameters such as the number of axes controlled through the isometric device, the level of stability and contol augmentation employed, and the demands of the primary task itself. The experimental data include certain objective measures of pilot workload: power spectral density analyses of control activity provide a measure of physical workload; mental workload is assessed using a secondary task based upon the Sternberg item recognition technique. This task, presented visually through the head-up display, is a memory scanning task requiring letter recognition and a verbal positive or negative response to each displayed letter. Reaction time and error data are collected and analyzed to assess the effects of the primary experimental variables. In addition to these workload measures, substantiating performance and pilot evaluation data are collected and analyzed.

DOUGLAS AIRCRAFT COMPANY

Ramona Gulick

ABSTRACT

VALIDATION OF PILOT WORKLOAD ESTIMATES UTILIZING IN-FLIGHT DATA

The purpose of this briefing is to describe the current capabilities of the Douglas Aircraft Company's workload evaluation techniques. The objectives of our workload evaluation studies are (1) to isolate design deficiencies and operational problems associated with specific hardware designs and functional arrangements of equipment during initial stages of development, (2) to verify that crew duty allocations and operational pcorcedures do not exceed the capabilities of individual crewmen, and (3) to evaluate crew composition.

The study approach includes the development of design mission scenarios and an in-depth evaluation of selected mission segments using task/timeline analytical techniques. Outputs of the crew workload evaluation program include: task loadings of individual crew members, equipment interface, and body channels. Outputs also include external vision availability and information processing time.

The derivation of task performance time estimates are based upon direct measurements recorded during mockup trials, in-flight analysis of video tapes, a reach envelope model, and the studies of other crew workload studies (Air Force, industry). Information processing time estimates were derived from studies/experiments cited in the literature and DAC laboratory evaluations of electronic flight display formats.

The accuracy of task time estimates has been validated through direct in-flight observations, including the FAA certification flight tests of commercial transport aircraft and micro-motion analysis of time-referenced video and audio recordings. Results of these studies showed a high correlation between estimated task time in actual in-flight observations.

THE FOCUS OF THIS BRIEFING IS: 1) TO DESCRIBE DOUGLAS AIRCRAFT's ANALYTICAL TECHNIQUES FOR FLIGHT CREW WORKLOAD EVALUATION, 2) TO EXPALIN THE METHODOLOGY EMPLOYED IN VALIDATING THE ANALYSIS, AND 3) TO SHOW THE APPLICATION OF THIS MODEL TO DETERMINE CREW COMPLEMENT AND DESIGN VERIFICATION.

THE VALIDATION OF THE ANALYTICAL TASK TIMES WAS DONE THROUGH IN-FLIGHT TESTING DURING THE DC-9 SUPER 80 CERTIFICATION PROCESS.

THIS ANALYSIS WAS APPLIED ON OUR RECENT C-X PROPOSAL TO THE AIR FORCE.

THE FOUR OBJECTIVES OF THE ANALYSIS WERE:

- VALIDATE THE ANALYTICAL METHODOLOGY THROUGH IN-FLIGHT DATA
- TO DETERMINE THE WORKLOAD IMPACT OF VARIOUS DESIGNS DURING THE DEVELOPMENT PROCESS
- VERIFY THAT OPERATOR TASK LOADINGS IMPOSED ON THE CREW BY THE STATION CONFIGURATIONS ARE NOT EXCESSIVE
- DETERMINE THE FEASIBILITY OF THE TWO FLIGHT CREW, ONE LOADMASTER CONCEPT.

IN THE STUDY APPROACH, IT WAS NECESSARY TO DEVELOP A REPRESENTATIVE FLIGHT SCENARIO TO EXERCISE THOSE FUNCTIONS CONSIDERED ESSENTIAL FOR DEMONSTRATING OPERABILITY FOR THE MINIMUM CREW.

AS A BASELINE FOR OUR ANALYTICAL TECHNIQUES, A MISSION WAS "FLOWN" WITH ALL SYSTEMS OPERATING NORMALLY AND THEN INTRODUCE INCREASINGLY DEGRADED MODES WERE INTRODUCED IN SUBSEQUENT SCENARIOS TO DETERMINE THEIR IMPACT ON THE WORKLOAD--AND LASTLY--TO DETERMINE THE ACCURACY OF THE TASK TIME ESTIMATES USED IN THE ANALYTICAL METHOD, ACTUAL IN-FLIGHT DATA WOULD BE ANALYZED FOR COMPARISON.

WORKLOAD IS DEFINED AS THE RATIO OF THE TOTAL TIME REQUIRED TO PERFORM A SERIES OF TASKS TO THE TIME AVAILABLE WITHIN THE CONSTRAINTS IMPOSED BY: THE MISSION, AIRCRAFT PERFORMANCE CHARACTERISTICS, AND CERTAIN EXTERNAL FACTORS SUCH AS ATC AND METEROLOGICAL CONDITIONS. THIS WORKLOAD IS EXPRESSED AS A NUMERICAL INDEX WHOSE MAXIMUM VALUE IS 100 (ALTHOUGH UNDER SOME SITUATIONS THE INDEX MAY EXCEED 100).

NUMBER 5

THIS FLOW DIAGRAM ILLUSTRATES THE PROCESS USED IN OBTAINING CREW WORKLOAD INDEX VALUES. THE MISSION SCENARIO AND TASK ANALYSIS WITH THE ASSOCIATED TIMES ARE CONSTRUCTED FROM THE PRELIMINARY AIR VEHICLE DESIGN. THE TIME AVAILABLE IS DETERMINED BY THE MISSION PROFILE, AIRCRAFT PERFORMANCE, AND ENVIRONMETAL CONDITIONS. THE TIME REQUIRED FOR EACH TASK IS DETERMINED BY USING A TASK/TIME DATA BANK THAT PROVIDES TIMES FOR REACHING AND ACTUATING VARIOUS TYPES OF CONTROLS AND FOR READING VARIOUS TYPES OF INSTRUMENTS IN RELATION TO A SPECIFIC TASK SEQUENCE. THIS METHODOLOGY WILL BE EXPLAINED LATER. THESE TIMES ARE BASED ON THE CREW STATION CONFIGURATION AND OPERATING PROCEDURES.

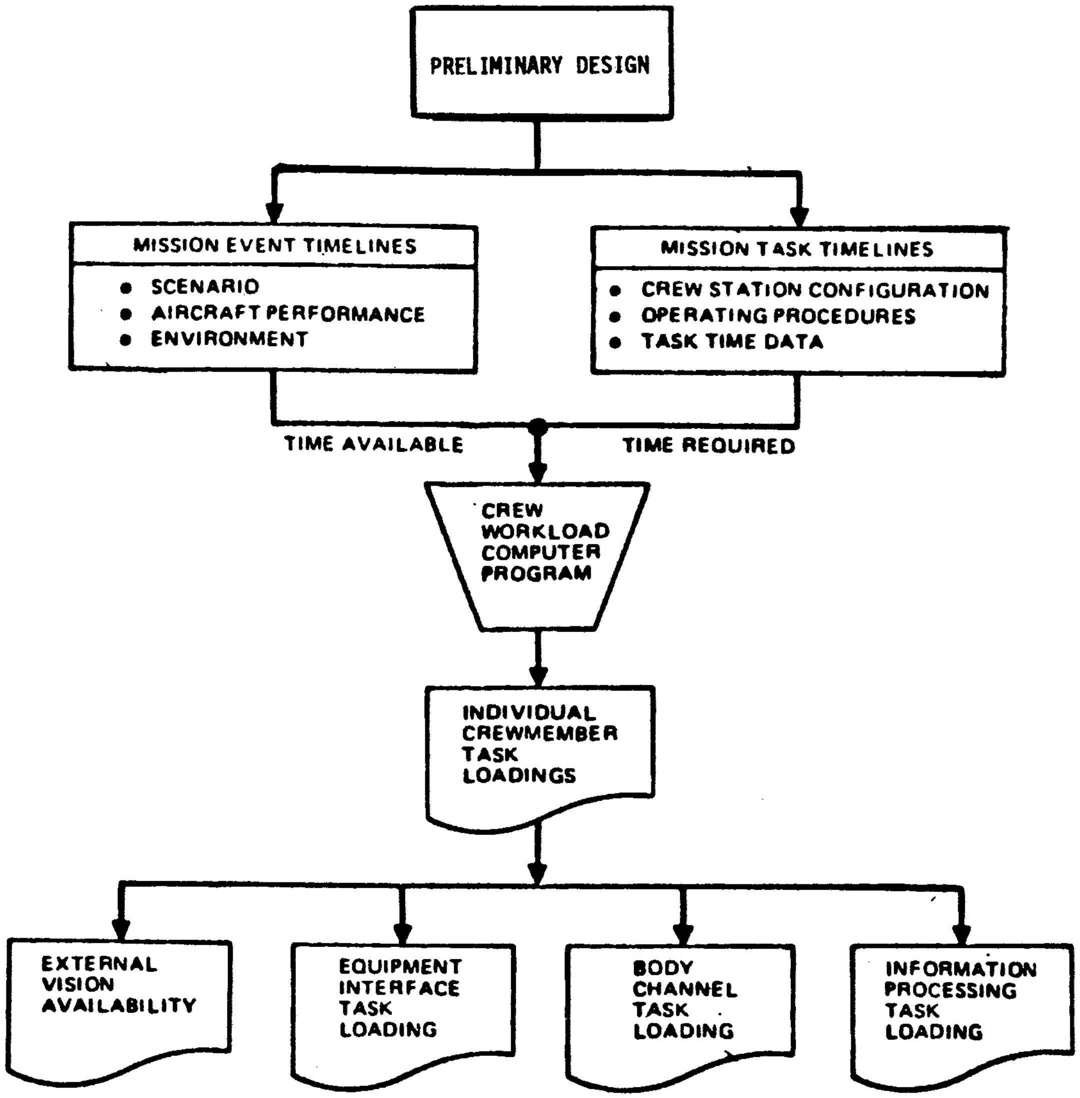
THE WORKLOAD COMPUTER PROGRAM CATEGORIZES AND ANALYZES THE DATA INTO VARIOUS OUTPUTS--WORKLOAD FOR EACH INDIVIDUAL CREW MEMBER, EXTERNAL VISION AVAILABILITY, TASK LOADINGS FOR EQUIPMENT INTERFACE, THE VARIOUS BODY CHANNELS, AND INFORMATION PROCESSING.

SEVERAL SOURCES WERE USED IN THE TASK TIME ESTIMATES.

- THE BASIC SUBTASK PERFORMANCE TIMES CAME FROM THE INDEX OF ELECTRONIC EQUIPMENT OPERABILITY DEVELOPED BY THE AMERICAN INSTITUTE FOR RESEARCH OR AIR DATA
- A DOUGLAS DEVELOPED EMPIRICAL FORMULA FOR ESTIMATING REACH TIMES AS A FUNCTION OF DISTANCE
- VERBAL COMMUNICATIONS WERE REPEATEDLY TIMED FOR A REALISTIC AVERAGE
- DIRECT ACTION TIME MEASUREMENTS WERE RECORDED DURING PROCEDURAL TRIALS IN THE CREW STATION DEVELOPMENT MOCKUP
- TIME REFERENCED VIDEO RECORDINGS ACQUIRED DURING PREVIOUS IN-FLIGHT MICRO-MOTION STUDIES WERE USED TO FURTHER ADD REALISTIC TIME ESTIMATES

TIMES FROM PREVIOUS LABORATORY INVESTIGATIONS RELATED TO VISUAL PERCEPTION DECISION MAKING ACTIVITIES WERE USED. AIR DATA PROVIDED INFORMATION PROCESSING TIMES IN MANY AREAS. TIMES DEVELOPED DURING EXPERIMENTAL TESTS CONDUCTED AT DOUGLAS IN SUPPORT OF THE ADVANCED ELECTRONIC DISPLAY FORMAT DEVELOPMENT WERE ALSO USED.

CREW STATION WORKLOAD EVALUATION PROCESS



- THE PROGRAM COMPUTES A WORKLOAD INDEX VALUE FOR EACH INDIVIDUAL CREWMEMBER AND REPRESENTS A COMPOSITE OF THE PHYSICAL AND COGNITIVE ACTIVITIES REQUIRED TO COMPLETE THE TASK SEQUENCE.

- WORKLOAD VALUES ASSOCIATED WITH SPECIFIC EQUIPMENT MAY BE TABULATED USING CATEGORY CODES SUCH AS ATA CHAPTER OR MILITARY SYSTEMS/SUBSYSTEM/SUBJECT NUMBERS TO ISOLATE OPERATIONAL EFFICIENTY OR INEFFICIENCY OF THAT PARTICULAR TYPE OF EQUIPMENT.

- THE TASK LOADING IMPOSED ON ONE INDIVIDUAL BODY CHANNEL MAY BE DETERMINED--RIGHT HAND, LEFT HAND, INTERNAL VISION, VERBAL/AURAL AND FEET.

- THE TIME REMAINING WHEN THE VISUAL MONITORING OF INTERNAL DISPLAYS AND CONTROLS IS NOT UTILIZED IS CONSIDERED EXTERNAL VISION AVAILABILITY OR TIME AVAILABLE FOR EXTERNAL ENVIRONMENTAL SCANS.

- ESTIMATED TIME VALUES WERE ALSO ASSIGNED FOR THOSE TASKS INVOLVING AN INFORMATION PROCESSING COMPONENT.

A DEDICATED FLIGHT TEST WAS MADE DUPLICATING AS CLOSE AS POSSIBLE THE SCENARIO DEVELOPED FOR THE ANALYTICAL STUDY. VIDEO AND AUDIO RECORDINGS WERE MADE AND ANALYZED THEN CORRELATED WITH THE ESTIMATED TIME VALUES. IN ADDITION, THE FAA CERTIFICATION FLIGHTS MADE ON THE EAST COAST TO DETERMINE CREW COMPLEMENT WERE RECORDED AND ANALYZED FOR FURTHER VALIDATION.

IN ORDER TO OBTAIN IN-FLIGHT VIDEO TAPE FOR THE DATA BASE VALIDATION, A DEDICATED TEST FLIGHT WAS ARRANGED DURING THE DC-9 SUPER 80 CERTIFICATION PROCESS. TWO CAMERAS--ONE FOR DAYTIME AND AN INFRARED FOR NIGHT--WERE MOUNTED ON THE COCKPIT BULKHEAD TO RECORD FLIGHT CREW ACTIVITY. THE FLIGHT PROFILE DUPLICATED AS CLOSE AS POSSIBLE THE SCENARIO USED IN THE TASK TIME ANALYSIS STUDY--A ROUND ROBIN LAX--LAX FLIGHT.

A TELEVISION RECEIVER WAS MOUNTED MIDWAY IN THE CABIN AREA. AIRSPEED, ALTITUDE, AND HEADING INDICATORS WERE INSTALLED BESIDE THE TV MONITOR WITH A BOX DISPLAYING DIGITAL READOUTS OF THE TWO NAVIGATION AND TWO VHF COMM FREQUENCIES SELECTED BY THE PILOTS. THE VHF RADIOS WERE TAPPED AND THE PILOTS EACH WORE A CHEST MICROPHONE FOR DISCRETE MONITORING OF AIR TO GROUND AND COCKPIT CONVERSATION. A TIME CODED VIDEO TAPE RECORDER SIMULTANEOUSLY TAPED BOTH THE VIDEO AND AUDIO PORTIONS OF THE FLIGHT FOR FUTURE ANALYSIS.

THE TECHNIQUE FOR ANALYZING THE VIDEO TAPE FOLLOWED TWO STEPS:

- FIRST THE TAPE WAS VIEWED AT NORMAL SPEED TO DETERMINE IF THE PROCEDURES AND DUTY ALLOCATIONS CORRESPONDED TO DATA BASE ESTIMATES. ALSO, IT WAS DURING THIS REVIEW THAT THE TASKS AND SUBTASKS COULD BE IDENTIFIED AS ONES WHICH COULD BE ACCURATELY TIMED.
- SECONDLY, THE TAPES WERE REVIEWED IN SLOW MOTION AND STOP/REVERSE/START MODES FOR THE MICRO-MOTION ANALYSIS.

NUMBER 13

AN EXAMPLE OF TWO TASKS COMPARISON IS SHOWN HERE. THE VIDEO TAPE "RETRACT GEAR" TASK IS TIMED FROM THE START OF THE PILOT'S CALL FOR "GEAR UP" THROUGH THE COPILOT'S MOTIONS AND THE RETURN OF HIS HAND TO REST. THE TASK FROM THE DATA BASE IS SUBDIVIDED INTO SUBTASKS.

NUMBER 14

IN FACT, A COMPARISON OF SIXTY-EIGHT DATA BASE AND THE CORRESPONDING VIDEO TAPE TASK TIMES INDICATE THAT THERE IS NO SYSTEMATIC TENDENCY TO OVERESTIMATE OR UNDERESTIMATE THE ACTUAL TIMES. IN FACT, THERE WAS A SLIGHT TENDENCY TO OVERESTIMATE THE ACTUAL TIMES RESULTING IN A PREDICTED WORKLOAD ON THE CONSERVATIVE SIDE.

NUMBER 15

STATISTICAL RESULTS PLOTTING THE ESTIMATED TASK TIMES VERSUS THE ACTUAL TASKS TIMES SHOW THE CORRELATION COEFFICIENT TO BE .848--VERY HIGH FOR A VALIDATION COEFFICIENT.

- OF THE SIXTY-EIGHT TASKS, THE AVERAGE TIMES FOR THE ACTUAL VALUES WAS 4.04 SECONDS ... FOR THE ESTIMATED 4.05 SECONDS
- NONE WERE OVER 13 SECONDS DURATION
- STANDARD DEVIATION OF ESTIMATED 2.73...ACTUAL 2.36

DC-9 SUPER 80

VIDEO TAPE		DATA BASE	
RETRACT GEAR		A. CAPTAIN CALLS "GEAR UP"	0.70
TIMED-FROM START OF CAPTAIN'S CALL FOR "GEAR UP" TO RETURN OF FIRST OFFICER'S HAND TO REST		B. FIRST OFFICER MOVES LEFT HAND TO GEAR LEVER	0.61
		C. MOVES LEVER TO UP POSITION	1.18
		D. RETURNS HAND TO REST	0.61
	2.4		3.10
TURN AUTOPILOT ON			
TIMED FROM START OF MOVEMENT OF CAPTAIN'S RIGHT HAND FROM WHEEL TO RETURN OF HAND TO REST		A. CAPTAIN MOVES RIGHT HAND TO AP SWITCH	0.61
		B. SETS SWITCH TO ON	0.48
		C. RETURNS HAND TO REST	0.63
	1.4		1.72

DC-9 SUPER 80

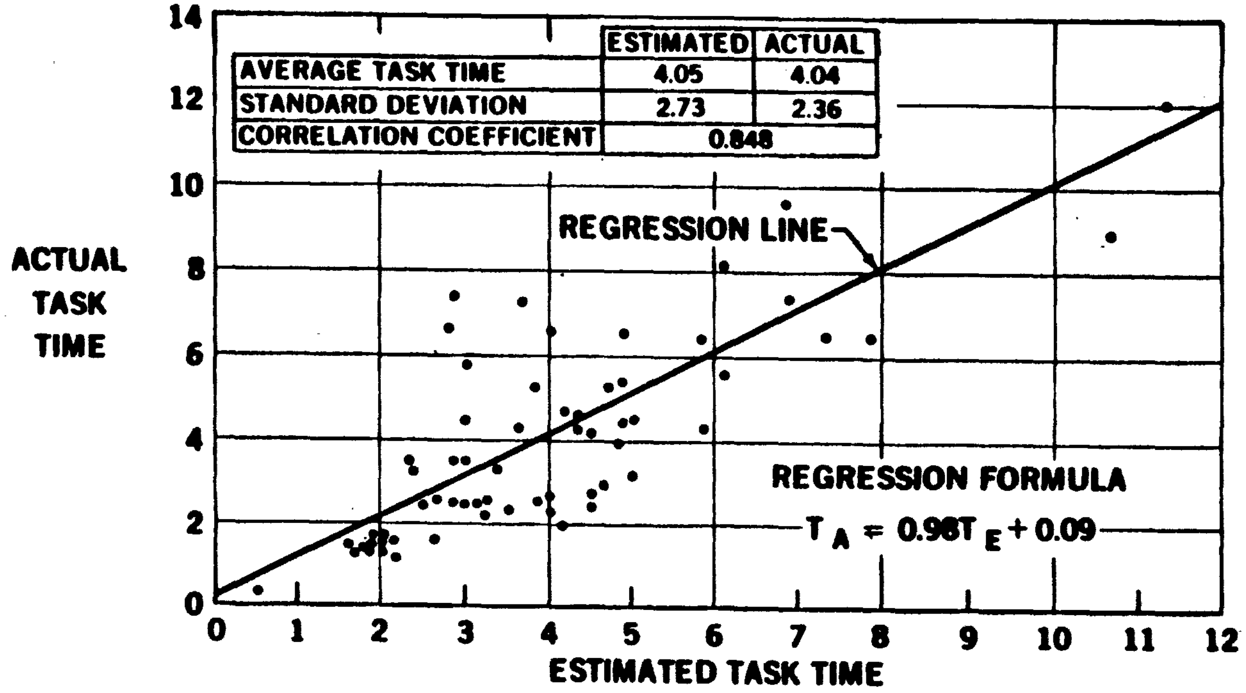
TASK TIME COMPARISON

VIDEO TAPE VERSUS DATA BASE (68 TASKS)

TASK TITLE	DATA BASE (SECONDS)	VIDEO TAPE (SECONDS)	DELTA (SECONDS)
1 RETRACT GEAR	3.1	2.4	—0.7
2 TURN ON AUTOPILOT	1.7	1.4	—0.3
3 INITIAL CLIMB AT 1000 FT	4.0	6.7	2.7
4 RETRACT FLAPS	4.2	2.0	—2.2
5 ADJUST THROTTLES	3.5	3.4	—0.1

DATA BASE VALIDATION

DC-9 SUPER 80



	ESTIMATED	ACTUAL
AVERAGE TASK TIME	4.05	4.04
STANDARD DEVIATION	2.73	2.36
CORRELATION COEFFICIENT	0.848	

$T_A = 0.98T_E + 0.09$

IN SUMMARY:

- THE MICRO-MOTION TECHNIQUE FOR VIDEO TAPE ANALYSIS PROVIDES FOR AN ACCURATE OBJECTIVE MEASUREMENT OF IN-FLIGHT TASK TIMES
- THE ESTIMATED DATA BASE TIMES CORRESPOND CLOSELY WITH THE MEASURED IN-FLIGHT VIDEO TAPE TASK TIMES
- ANALYTICAL MODEL PROVIDES AN ACCURATE INDEX TO PREDICT IN-FLIGHT WORKLOAD LEVELS

THE PREVIOUS SLIDES SHOWED THE METHODOLOGY AND VALIDATION RESULTS FROM THE DC-9 SUPER 80 CERTIFICATION PROGRAM.

THIS SAME ANALYTICAL MODEL WAS APPLIED IN DOUGLAS' RESPONSE TO THE AIR FORCES' C-X RFP IN WHICH THE FOLLOWING REQUIREMENTS WERE ISSUED.

- WORKLOAD REDUCTIONS SHOULD BE ACCOMPLISHED BY TASK ANALYSIS
- VERIFY THE INTEGRATION WITHIN AND BETWEEN FLIGHT DECK AND CARGO COMPARTMENT

THE AIRDROP MISSION SEGMENT WAS SELECTED FOR THE INITIAL CREW WORKLOAD EVALUATION FOR THREE BASIC REASONS:

- IT PLACES SUBSTANTIAL DEMANDS ON THREE FUNCTIONS IDENTIFIED PREVIOUSLY AS "MISSION CRITICAL"

  NAVIGATION
  STATION KEEPING
  AERIAL DELIVERY

- IT PROVIDES AN OPPORTUNITY TO EVALUATE WORKLOAD IMPOSED ON ALL THREE CREW MEMBERS--PILOT, COPILOT, AND LOADMASTER.

- THE AIRDROP MISSION IS ALSO CONSIDERED BY CURRENT MAC CREWS TO BE A HIGH WORKLOAD MISSION.

NUMBER 19

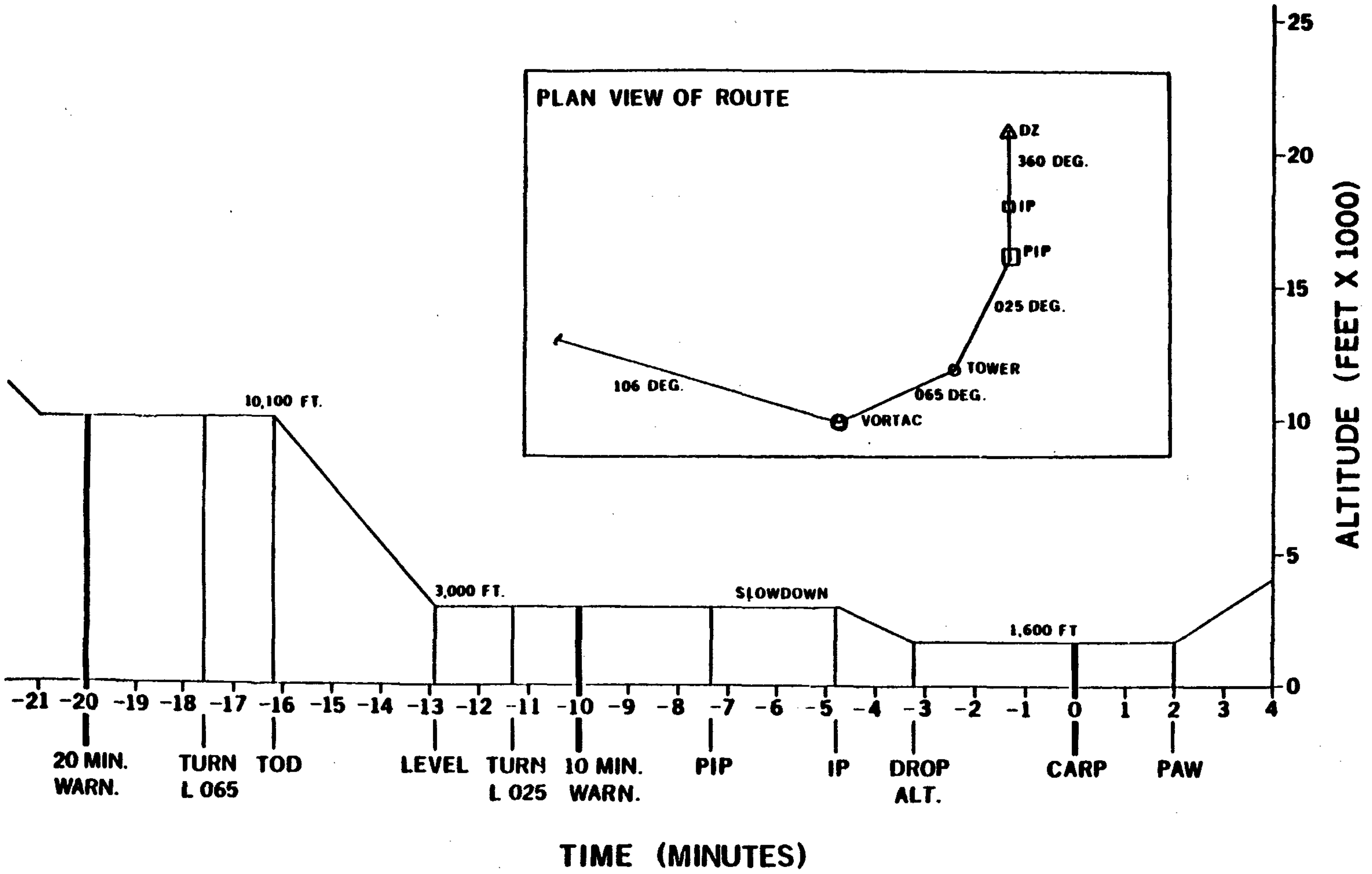
IN ORDER TO INSURE THAT THE TASK/TIMELINE ACCURATELY REFLECTS THE OPERATING ENVELOPE OF THE AIRCRAFT, A DETAILED ALTITUDE/TIME PROFILE WAS DEVELOPED BASED ON THE ANTICIPATED PERFORMANCE CHARACTERISTICS OF THE AIRCRAFT AND MILITARY PROCEDURES FOR A REPRESENTATIVE AIRDROP.

NUMBER 20

CREW DUTY ALLOCATIONS WERE CALCULATED TO DISTRIBUTE TASKS EQUALLY WHILE RETAINING COMMAND/DECISION RESPONSIBILITIES:

- PRIMARY PILOT DUTIES (SECONDARY COPILOT DUTIES)
- PRIMARY COPILOT DUTIES (SECONDARY PILOT DUTIES)
- PRIMARY LOADMASTER DUTIES

C-X AIRDROP MISSION SEGMENT
ALTITUDE - TIME PROFILE



CREW DUTY ALLOCATIONS

PRIMARY PILOT DUTIES (SECONDARY CO-PILOT DUTIES)	PRIMARY CO-PILOT DUTIES (SECONDARY PILOT DUTIES)	PRIMARY LOADMASTER DUTIES
SELECT/MONITOR CONTROL SYSTEM MODES	COMMUNICATIONS MANAGEMENT	CONTROL AND MONITOR ADS
SELECT/MONITOR NAVIGATION SYSTEM MODES (NAV 1)	SELECT/MONITOR NAVIGATION SYSTEM MODES (NAV 2)	INTERCOMPARTMENT COMMUNICATION
MONITOR CAUTION & WARNING SYSTEM	MANAGE AIRCRAFT SUBSYSTEMS	HANDLE SAFETY EQUIPMENT
CALL/BRIEF FOR CHECKLISTS	MONITOR CAUTION & WARNING SYSTEM	EXECUTE CHECKLISTS
	EXECUTE CHECKLISTS	

NUMBER 21

THIS TABLE PROVIDES A LISTING OF THE INPUT PARAMETERS NECESSARY TO CONSTRUCT A SAMPLE TASK--OPEN CARGO DOOR AND RAMP. THIS IS ONLY ONE OF THE TASKS WHICH OCCURS DURING THE MILESTONE BOUNDED BY THE PIP (OR PRE-INITIAL POINT) AND THE IP (INITIAL POINT). THE TASK IS FURTHER SUBDIVIDED INTO SUBTASKS WHICH ARE INDIVIDUAL AIRCREW ACTIONS REQUIRED TO PERFORM THE TASK.

NUMBER 22

INSPECTION OF THE GRAPHIC OUTPUT OF TASK LOADING BY MILESTONE REVEALS A GENERALLY LOW WORKLOAD FOR BOTH THE PILOT AND COPILOT. THE MAXIMUM WORKLOAD INDEX FOR ANY MILESTONE DOES NOT EXCEED SIXTY.

NUMBER 23

THE WORKLOAD PROFILES PRESENTED PREVIOUSLY ASSUME THAT ONLY THE CONTROL STICK STEERING MODE OF THE AUTOFLIGHT CONTROL SYSTEM (AFCS) IS UTILIZED. A SECOND ANALYSIS WAS CONDUCTED TO ASSES THE IMPACT OF USING THE FULL AFCS ON THE WORKLOAD LEVELS OF THE PILOT.

RESULTS SHOW A SIGNIFICANT DECREASE IN WORKLOAD LEVELS RESULTING FROM FULL USE OF THE AUTOFLIGHT CONTROL SYSTEM AND DEMONSTRATE THE COMPARISON CAPABILITIES OF THE PROGRAM.

INPUT PARAMETERS FOR A REPRESENTATIVE TASK
AIRDROP MISSION SEGMENT

TASK: OPEN CARGO DOOR AND RAMP

MILESTONES	TASK	SUBTASK	COMMUNICATIONS	CREW MEM.	EQUIP. INTER-FACE	BODY CHAN.	SUBTASK TIME (SEC.)	INFO. PROCES. TIME (SEC.)
START PIP	OPEN DOOR & RAMP	VERIFY DOOR & RAMP CLEAR		LM	113L	E	0.25	0.10
		MOVE HAND TO MIKE PTT BUTTON		LM	235L	R	0.67	0.00
END IP		PRESS PTT BUTTON		LM	235L	R	0.47	0.10
		REPORT 'DOOR AND RAMP...'	'DOOR & RAMP CLEAR, LOADMASTER'	LM	2350	VR	2.80	0.25
		HEAR 'DOOR & RAMP' REPORT		P	2350	V	2.80	0.25
		HEAR 'DOOR & RAMP' REPORT		CP	2350	V	2.80	0.25
		CALL 'DOOR & RAMP OPEN'	'DOOR & RAMP OPEN'	P	2350	V	1.20	0.25
		HEAR 'DOOR & RAMP OPEN'		CP	2350	V	1.20	0.25
		MOVE HAND TO ADS DOOR & RAMP SW		CP	9412	EL	0.65	0.00
		PULL SW OUT OF DETENT		CP	9412	EL	0.25	0.10
		SET SW TO OPEN POSITION		CP	9412	EL	0.71	0.40
		RETURN HAND TO REST		CP	9412	L	0.65	0.00

CREWMEMBER CODES

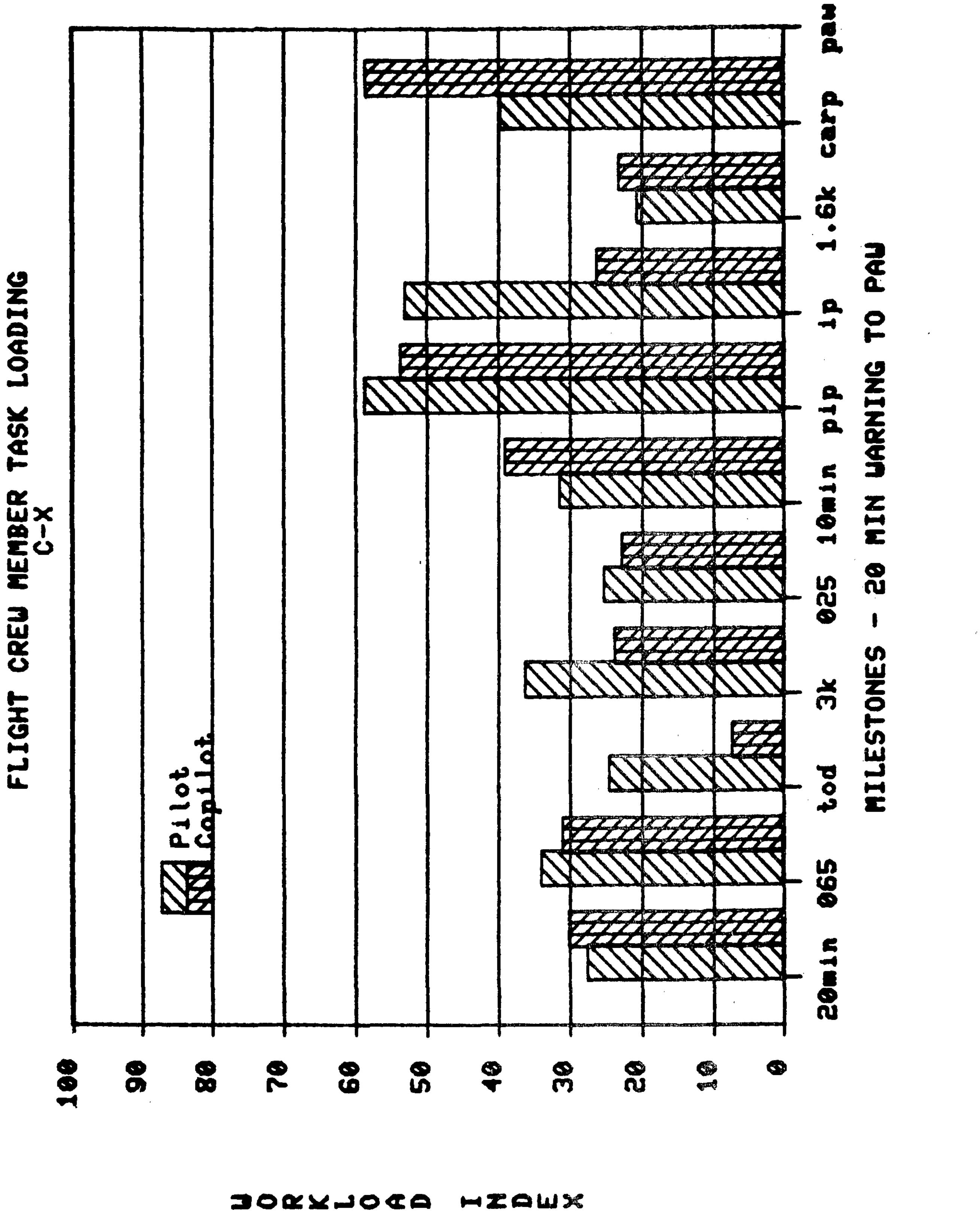
P = PILOT
CP = COPILOT
LM = LOADMASTER

BODY CHANNEL CODES

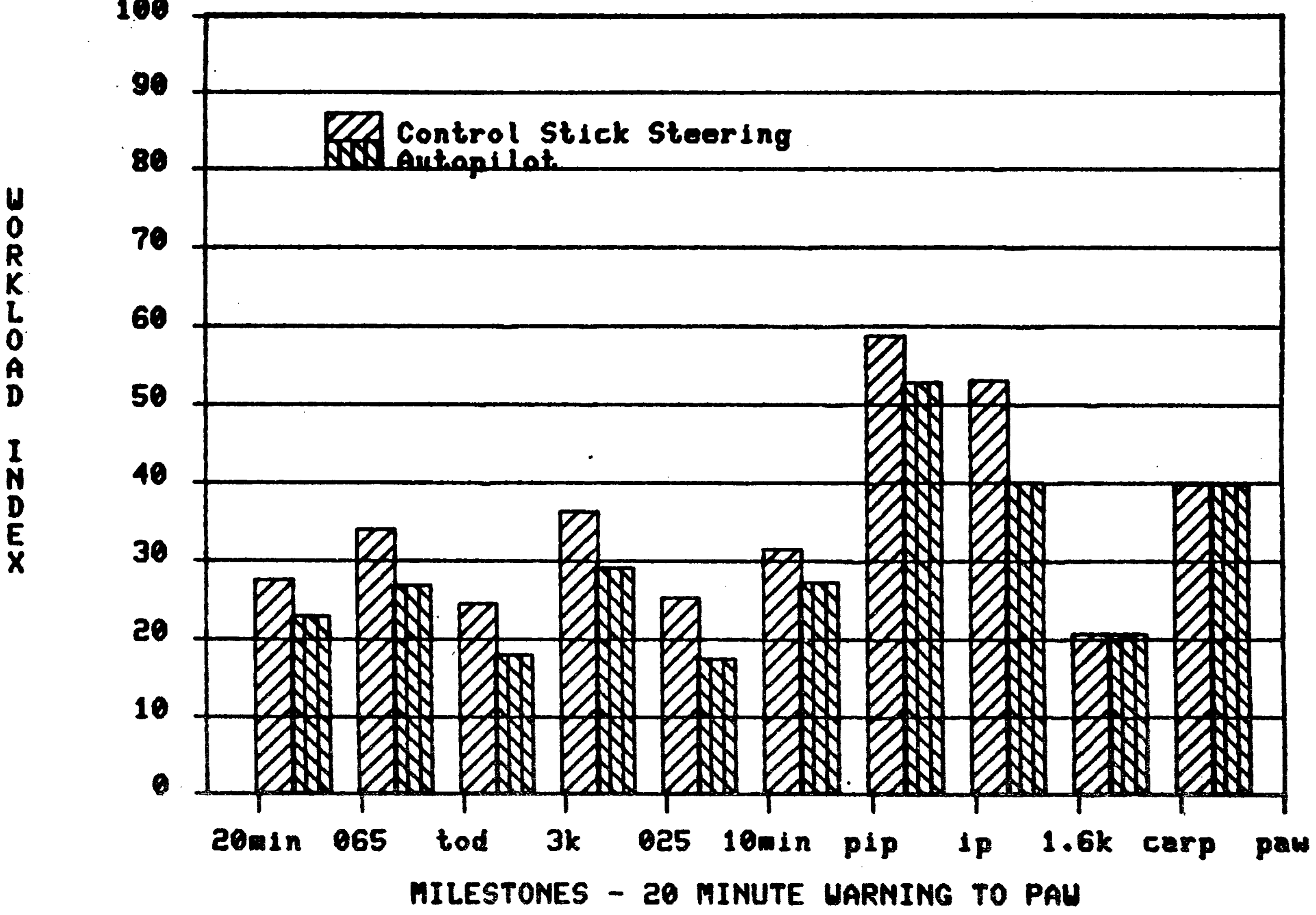
V = VERBAL/AURAL
E = INTERNAL VISION
L = LEFT HAND
R = RIGHT HAND

EQUIPMENT DESIGNATIONS

113L = VERIFICATION
235L = PTT SW
23501C = COMMUNICATIONS - INTERCOMM.
9412 = ADS DOOR AND RAMP SW


FLIGHT CREW MEMBER TASK LOADING
C-X
Pilot
Copilot
WORKLOAD INDEX
100
90
80
70
60
50
40
30
20
10
0
20min
065
tod
3k
025
10min
pip
ip
1.6k
carp
paw
MILESTONES - 20 MIN WARNING TO PAW

CSS VERSUS AP MODE COMPARISON
Pilot



NUMBER 24

THIS TABLE SUMMERIZES THE OVERALL TASK LOADINGS ASSOCIATED WITH EACH FUNCTIONAL EQUIPMENT CATEGORY AND HIGHLIGHTS POSSIBLE AREAS FOR AUTOMATION OF CERTAIN SYSTEMS IF AN EXCESSIVE WORKLOAD EXISTS.

THE RESULTS OF THIS ANALYSIS REVEAL AN EVEN DISTRIBUTION OF TASKS WITH THE WORKLOAD INDEX FOR THE VARIOUS CREWMEMBERS WELL WITHIN AN ACCEPTABLE RANGE.

THE CREW STATION CONFIGURATION ALLOWED FOR FLEXIBILITY IN DUTY ALLOCATIONS SO AS TO INSURE MISSION ACCOMPLISHMENT WITH THE PROPOSED MINIMUM CREW COMPLEMENT.

THEREFORE, WE FEEL THROUGH THE VALIDATION OF DOUGLAS' WORKLOAD MODEL WE CAN, WITH SOME CERTAINTY, PREDICT IN-FLIGHT WORKLOAD LEVELS.

EQUIPMENT UTILIZATION

FLIGHT STATION

SYSTEMS	WORKLOAD INDEX	
	PILOT	CO-PILOT
FLIGHT CONTROLS	10.57	2.52
NAVIGATION	5.44	8.92
COMMUNICATION	8.55	9.76
SYSTEM MANAGEMENT	0.24	2.85
STATION KEEPING	2.83	1.55
AERIAL DELIVERY	3.03	6.46
DOCUMENTATION	0.00	0.00

CARGO COMPARTMENT

SYSTEMS	WORKLOAD INDEX
EXTRACTION	16.89
RESTRAINTS	6.19
DOORS AND RAMP	1.93
SAFETY EQUIPMENT	4.13
COMMUNICATION	8.12
DOCUMENTATION	0.00
CONTROLS & DISPLAYS	1.58

THE ELUSIVE GOAL OF MEASURING PILOT WORKLOAD IN GENERAL AVIATION

Earl S. Stein John Fabry Bruce Rosenberg

Federal Aviation Administration Technical Center

The magnetic attraction of people to aviation has existed since the beginning of manned flight. This is particularly true in general aviation where the majority of pilots are not only unpaid but, in fact, pay their own expenses. For these individuals and others in aviation, potential changes in aircraft cockpits must clearly demonstrate advantages in excess of costs.

Given a highly complex, multidimensional stimulus/task environment, how does one determine pilot workload and performance? These two global labels are the basic elements of what could be called overall "pilot effectiveness." It has been very apparent from the literature in aviation and human factors psychology that usable systematic measures are not currently available for general aviation. The Federal Aviation Administration (FAA) Technical Center in Atlantic City, New Jersey, has a requirement to evaluate the impact of both new and altered cockpit systems. The human impact of these changes must be quantified in terms of "pilot effectiveness."

A detailed examination of the literature in human workload has led to certain conclusions. No one questions the complexity or multidimensional nature of workload. However, models have been largely design-specific with each experimenter tailoring his/her theoretical model to meet his/her immediate needs and many models stress the atomization of workload into more "basic elements," such as mental and physical components. Measurement methods have included physiological techniques through performance indices and psychological approaches. The most popular method by far has been the postflight questionnaire. This technique is easy to administer but relies heavily on pilot memory for accuracy. Given all the effort expended, there has been a surprising lack of success in aviation workload measurement, and it has not been well linked to any clear definition of task difficulty.

The literature in the area of pilot performance dates well back into World War II and beyond. It has largely been focused on the training environment where cost-benefit assessment is constantly emphasized. The most frequently used technique has been instructor/pilot ratings. While these can be made reasonably objective and valid, often they are not, and depending heavily on face validity, which means essentially how good they look to the people who must use them. Individual instructor/pilot interests and biases can reduce measurement reliability considerably. Such techniques may serve a valuable need as training diagnostic indicators, but they lack precision to identify what may be subtle changes in pilot behavior induced by cockpit systems alterations. In the past few years, there has been a new trend, primarily in military aviation, to employ a computerized data collection system to evaluate performance. Automated performance measurement (APM) allows for the collection of aircraft state and control input variables at a very high sampling rate. Data can be accurately and objectively recorded. The method assumes that inferences can be be made about pilot performance based on how the aircraft is flying and on the nature and intensity of control inputs.

The FAA Technical Center has developed its own workload and performance models. While these models are not completely atheoretical, pragmatism rather than theory has been emphasized. The workload model assumes that if you ask a pilot how hard he/she is working, he/she will be able to tell you. This is provided that the question is properly timed and formatted. The model further assumes that pilots' responses will weigh the input of all stressors which currently effect his/her behavior. No attempt will be made to force the pilot to atomize workload into multiple catagories since this would make the workload response requirement more intrusive. The final assumption made by the model is that the best time to measure workload is during the flight itself.

The pilot performance model stresses automated performance measurement. Performance, unlike workload, can be observed directly and can be measured unobtrusively. Here an atomistic approach will be taken. Subject matter experts have identified eight segments of a normal general aviation flight. These include: 1. takeoff, 2. climb, 3. en route, 4. holding, 5. descent, 6. initial approach, 7. final approach, and 8. missed approach. Each segment contains specific variables which the subject matter experts felt were critical for evaluating performance. For example, the takeoff segment contains heading, airspeed, manifold pressure, rpm, pitch, and bank angles. Each variable has an assigned ideal value and acceptable limits around that value. The model assumes that the results for all variables can be summed to provide a segment total score and that segment scores can be pooled for a total flight performance score, which will be called the pilot proficiency index or PPI. Since the construction of the PPI has been analytical, empirical verification will be essential before it can be put to any practical use.

Two workload studies have been completed already. The first study employed a nonflying, two-axis, critical tracking task. Twenty-four people, of whom 12 were pilots, were asked to maintain a point of light in the center of a cathode-ray tube (CRT) display using a joystick control. This task has been compared to attempting to balance a ball on the end of a broomstick. The further one strays from the center, the more difficult it becomes. The driving force behind the display was an analog computer which could alter the difficulty level along a continuum. Each participant's maximum ability was measured and then difficulty levels were tailored as proportions of their own maximums. These proportions included four levels: .25, .50, .75, and 1.00. Trials were run in a counterbalanced order. During each 4 1/2 minute trial, a query tone was sounded every minute. The participant was asked to press a button from 1 (very easy) to 10 (very hard) which best described how hard he/she was working. Results indicated that participants were willing and able to make the workload evaluations during primary task performance. Their workload values were directly related to the difficulty level assigned for any given trial. Response latency, however, was not related to difficulty.

A second workload study employed a Singer-Link General Aviation Trainer (GAT-2) which simulated a light twin cabin class aircraft (Cessna 421). Three qualitatively different levels of flight difficulty were assigned. Flight difficulty was controlled by manipulating clearance complexity, air traffic control complexity, air turbulence, and an inflight emergency. Twelve pilots each flew three flights in a counterbalanced order. They were instructed to evaluate their inflight workload using a 10 pushbutton scale mounted in a box below the throttles. During flight, their responses were requested every minute using both a tone and a light. The tone lasted only one-fifth of a second while the light (a small LED on the switchbox) stayed on until a button was pressed. Results indicated that workload responses increased significantly in direct relationship to the difficulty level. Response delays also increased with difficulty but only between the easiest and the moderately difficult flights. There was no difference between the moderate and most difficult flights in terms of response delay. One explanation for this is the possibility that as difficulty increased, the participant's attention became more focused on workload and on their evaluations of it. The second study clearly demonstrated that pilots were able to make workload discriminations during simulated flight. The results of postflight questionnaire indicated that pilots did not find using the pushbuttons to be overly obtrusive.

While workload work will continue, performance research is just beginning at the FAA Technical Center. An initial test of the PPI concept will involve flying two diverse groups in the

Center's GAT-2. These groups sill include high-time, professional pilots and low-time, just qualified instrument/multiengine pilots. Assuming that the PPI can discriminate between these two groups, additional tests will be accomplished to validate the method. While in the short run, APM will be confined to simulation, long-term possibilities could include airborne flight tests using state-of-the-art microcomputer techniques for data collection.

An airborne workload data collection system has already been constructed in prototype. This will be evaluated based on the availability of flight time. It is likely that it will be used in conjunction with other projects.

Techniques for measuring workload and performance in general aviation are under active development. The goal is to establish the tools necessary in order to provide timely and accurate information concerning the effects of systems changes on pilot behavior. Only through active, empirical research can such tools be developed.

SUBJECTIVE WORKLOAD ASSESSMENT TECHNIQUE

Gary B. Reid, F. Thomas Eggemeier
and Clark A. Shingledecker

Subjective assessment techniques are often employed in flight testing and operational test and evaluation. The acceptability of these measures is best demonstrated by the widespread use and utility of the Cooper-Harper scale for assessing aircraft handling qualities.

We in the Workload and Ergonomics Branch of the Air Force Aerospace Medical Research Laboratory are engaged in a program of research to develop a comprehensive battery of tests for assessing pilot workload. A considerable part of this program is devoted to the development and validation of physiological measures, behavioral measures, and systems performance measures. Even though the primary emphasis of our program is not the development of subject measures we felt that, in order for our test battery to be complete, we needed a systematically developed and validated subjective measure that is generally applicable and widely accepted.

A thorough literature review of subjective measures revealed that precisely what we wanted does not currently exist. While subjective measures are frequently used in workload assessment, they usually are designed for a specific application. The other category of readily available measures includes measures like the Cooper-Harper and SOMA scales. These scales assess handling qualities and system's operability respectively and, as such, are focused on systems evaluation with workload as a component.

The measure we desire should be designed specifically to assess workload. Within the context of our battery this measure is conceived as being less precise than some of the other proposed measures (e.g. cortical evoked response) but should be precise and sensitive enough to quantify the existence of high workload. Also, this scale should be simple to administer so that applications such as operational test and flight test are possible. Since none of the currently existing scales that we reviewed possessed these attributes, we initiated an effort to develop our own subjective measure which we have named the Subjective Workload Assessment Technique (SWAT).

To develop SWAT we defined workload as being primarily composed of three dimensions: (1) Time load, (2) Mental effort load, and (3) Psychological stress load. This multidimensional definition of workload is consistent with a large body of current workload research (Johannsen, et al, 1979; Senders, 1979; Williges and Wierwille, 1979; Johns, 1973; White, 1971; Sheridan and Simpson, 1979). In particular, this framework and the definitions of the three dimensions are adaptations of work performed by Sheridan and Simpson (1979) in development of a Cooper-Harper type of category scale for workload. Each of the three dimensions has three levels corresponding roughly to high, medium, and low loading. Precise definitions and a more in-depth discussion of SWAT development can be found elsewhere (Reid, et al, 1981a; 1981b).

These definitions of three levels of each of three dimensions are combined through a mathematical procedure known as conjoint measurement.

This portion of the SWAT procedure we have named Scale Development. During scale development information is obtained from subjects regarding the way the three dimensions theoretically combine to produce workload. This information is collected by having subjects order all possible combinations of the definitions. This means all possible combinations of the three levels of the three dimensions or 27 combinations. The ordering is a separate distinct process which is carried out, preferably, before starting an experiment. The data obtained is then tested for agreement among subjects. If sufficient agreement exists, say a Kendall's coefficient of condordance greater than .75, then the remainder of the procedure is based upon group data. It is possible, however, to develop an independent scale for each subject in the event that there is low agreement.

After deciding whether to obtain one scale to represent the group or a scale for each subject the data is tested by a procedure called conjoint measurement (Krantz, 1964; Krantz and Tversky, 1971; Luce and Tukey, 1964). This procedure tests the ordering data to determine the combination rule used by the subjects. Several rules or models are possible (e.g. additive, distributive, dual distributive) but all applications of SWAT have resulted in an additive model.

Once the model is defined, the data is transformed via an iterative procedure known as conjoint scaling. The transformation fits the data to the defined model and maintains the order inherent in the orginal data. The result is a undimensional interval level scale.

The event scoring phase of SWAT is the process of obtaining workload ratings within the context of the investiators experiment or evaluation. This is done by having subjects provide ratings of the event in terms of the levels of the three dimensions inherent in the current situation. For example, a subject might assign a 3 for high time load, a 2 for moderate mental effort, and a 2 for psychological stress as his ratings for coping with a particular in-flight emergency. These ratings then might correspond to a value of 125 on the previously developed scale and this value is the workload level associated with the event.

The initial step in the systematic development of SWAT was to see if subjects could perform the required orderings, and if so, what level of agreement exists between subjects and within subjects. The first question has been addressed by having eight groups of subjects perform the ordering procedure. The entire procedure normally takes approximately forty-five minutes. Kendall's coefficients of concordance associated with these groups ranged from .79 to .87 indicating statistically reliable agreement within the groups. The coefficient of concordance obtained from the pooled data of all subjects is .80 indicating high between group agreement.

Reliability checks are made for individual subjects by follow-up administration of a set of paired comparisons. Subjects are presented with the set of three descriptions describing two workload conditions and are asked to indicate which set represents highest workload. Subjects are presented nine of these pairs where three represent very fine discriminations, three represent moderate discrimination and three represent discrimination of situations clearly having different degrees of workload. Using this technique, subjects have been found to be approximately 80% reliable when checked about one month of scale development and approximately 75% reliable after a lapse of about four months.

In order to assess validity of SWAT an experiment was conducted where subjects were presented with eight aircraft communications tasks while performing a critically unstable tracking task. The communications tasks had previously been scaled using an information theoretic analysis (Shingledecker et al, 1980) so that the eight task were known a priori to possess varying levels of task demand. Additionally, two levels of task difficulty were use. Thus, sixteen dual task conditions were presented. The results of this experiment (Reid, et al, 1981a) demonstrated that SWAT was able to discriminate between the two levels of primary task as well as between the eight levels of the secondary task.

Additional validation data has been obtained through implementing SWAT as a workload assessment instrument in systems tests. The most noteworthy of these applications is the Advanced Medium Range Air-to-Air Missile (AMRAAM) Operational Utility Evaluation (OUE). The OUE was a large scale AIMVAL/ACEVAL type of exercise conducted by the Air Force Test and Evaluation Center (AFTEC) using the MAC AIR combat simultor at McDonnell Douglas Corporation, St Louis, MO. In the course of this evaluation we collected over 12,000 event scores from a pool of fifty-six pilots. A review of the frequency distribution of these scorings reveals that the whole range of the twenty-seven point scale is used. Additional information from this test showed workload differences in places and in directions that can be logically defended. The exact findings cannot be discussed here due to the classification of the data. However, post test questionnaires and pilot interviews were administered and the subjective opinions of the test participants support the findings obtained from statistical analysis of the event data.

Additional applications of SWAT are currently in the planning stages. An AFAMRL/AFWAL program (RAM/ACE) to investigate potential cockpit enhancements is employing a modified projective procedure using SWAT. The AFTI-16 program as well as the single-seat night attack study will provide the first flight test applications.

Refinement of the technique will continue as we gain experience in a wide variety of applications in order to provide a general workload measure. Data is being collected regarding SWATs sensitivity, validity and reliability in order to aid potential users apply the technique to their own specific application.

Johannsen, G., Moray, N., Pew, R., Rasmussen, J., Sanders, A., and Wickens, C. Final report of the experimental psychology group. In: N. Moray, (Ed.), *Mental Workload: Its Theory and Measurement*. New York: Plenum Press, 1979.

Johns, D. W. (1973). A Concept of Operator Workload in Manual Vehicle Operations. Meckenheim, Germany: Forschungsinstitut fur Anthropotechnik, Report No. 14, 1973.

Krantz, D. H. Conjoint measurement: the Luce-Tukey aziomatization and same extensions. *Journal of Mathematical Psychology*, 1964, *1*, 248-277.

Krantz, D. H. and Tversky, A. Conjoint measurement analysis of composition rules in psychology. Psychological Review, 1971, 78, 151-169.

Luce, R. D. and Tukey, J. W. Simultaneous conjoint measurement: A new type of fundamental measurement. Journal of Mathematical Psychology: 1964, 1, 1-27.

Reid, G. B., Shingledecker, C. A. and Eggemeier, F. T. Application of Conjoint Measurement to Workload Scale Development. Proceedings of the 1981 Human Factors Society Annual Meeting, October 1981.

Reid, G. B., Shingledecker, C. A., Nygren, T. E. and Eggemeier, F. T. The Development of Multidimensional Subjective Measures of Workload. Proceedings of the 1981 International Conference on Cybernetics and Society, October, 1981.

Sanders, A. F. Some remarks on mental load. In: N. Moray (ed.), Mental Workload: Its Theory and Measurement. New York: Plenum Press, 1979.

Sheridan, T. B. and Simpson, R. W. Toward the definition and measurement of the mental workload of transportation pilots. Cambridge, Mass. Massachusetts Institute of Technology Flight Transportation Laboratory Report, FTL Report R79-4, January 1979.

Shingledecker, C. A., Crabtree, M. S., Simons, J. C., Courtright, J. F., O'Donnell, R. D. Subsidiary radio communications tasks for workload assessment in R&D simulations: 1. task development and workload scaling Air Force Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, AFAMRL-TR-80-126, 1980.

White, R. T. Task analysis methods: Review and development of technique for analyzing mental workload in multiple-task situations St. Louis, Missouri: McDonnell Douglas Corporation, MDC J5291, September 1971.

Williges, R. C. and Wierwille, W. W. Behavioral measures of aircrew mental workload. Human Factors, 1979, 21, 549-574.

ASSESSMENT OF RELIABILITY AND VALIDITY
FOR THE
SYSTEMS OPERABILITY MEASUREMENT ALGORITHM (SOMA)

By

Kendrick N. Williams, Ph.D.
Human Factors Engineering Branch
Code 1226
Pacific Missile Test Center
Point Mugu, CA 93042

Assessment of Reliability and Validity for the Systems Operability Measurement Algorithm (SOMA)

The Systems Operability Measurement Algorithm (SOMA) (reference 1) was developed as a system evaluation methodology to provide decision makers with information about system adequacy and design alternatives within the context of mission needs and requirements. During the past year a research effort has been directed toward evaluation of SOMA in terms of its reliability and validity.

The SOMA model has evolved directly from multi-attribute utility (MAU) theory (reference 2). MAU is a Bayesian oriented decision making paradigm (reference 3) which processes information according to a set of specific rules.

There are three rules of MAU which are particularly important to SOMA. First, the basic structural principle in MAU is hierarchical decomposition. This means that the evaluation problem is broken down from general to specific components. The model provides the structure and rules necessary to investigate and integrate the interrelationships of all components. Second the MAU model involves the definition of utility (effective operability). Optimum evaluation of alternatives is dependent upon the selection of a single criterion. Multidimensional outcomes must be transformed into a single figure of merit such as utility system worth or operational effectiveness. Third, the MAU model utilizes the scaling of the selected criterion.

These three aspects of MAU were addressed in the development of SOMA. As part of SOMA development, task analysis methodology was used to establish the decomposition hierarchy and to provide the basis for the utilization of operability measures and scaling theory. These scaling methodologies are employed to develop the rating scales for the criticality (C), pilot workload (PW) and technical effectiveness (TE) dimensions of operability, and to provide the basis for integrating PW and TE into a single intervalized effective operability (EO) dimension.

As illustrated in Figure 1, this approach for system evaluation integrates the technologies of task analysis, multi-attribute utility theory, and scaling theory into one functional algorithm. SOMA systematically structures operator tasks (Figure 2) in accordance with mission needs hierarchy (Figure 3) and then provides for assessment of system operability through application of a linear conjoint measurement scaling model (Figure 4). All assessments are then integrated by the rules established through MAU theory. The output from SOMA is quantitative information about the operability of an entire weapons system (i.e., fighter or attack aircraft), the operability of specific systems (such as radar, communication or navigation) and, finally, the operability of each task performed during a mission phase. In short, SOMA measures subsystem and/or system goal attainment.

- -

Insert Figures 1-4 About Here

- -

In order to demonstrate the validity and reliability of the overall SOMA technique, the reliability and validity of each component part must first be demonstrated. Although SOMA is currently being used as an operational test and evaluation tool by the U.S. Air Force, it is appropriate to examine the psychometric characteristics of SOMA. Examination of these characteristics will provide the basis for further operational refinement of SOMA. It is thus necessary to:

- Compute the test-retest reliability (pilot rating of tasks over time) demonstrated for the criticality (C), technical effectiveness (TE), pilot workload (PW) and effective operability (EO) scales which compose the SOMA model.

- Compute the interrater reliability (pilot concurrence on scale meaning and agreement) for the C, TE, PW and EO scales of the SOMA model.

- Compute internal consistency reliability of pilot responses for the MAU's task inventory.

- Assess the face validity of SOMA, i.e., determine whether users agree that the instrument and the scales adequately measure tasks important to the mission.

- Assess the construct validity of SOMA, i.e., empirically examine the hypothesized relationships among the structural components of the SOMA model.

In order to address these reliability and validity questions, A-7 and F-16 pilot inventory responses were analyzed. Questions of reliability are addressed by employing a variety of standard reliability analyses and estimates of concordance. Face validity is addressed by a recapitulation of both the development of the hierarchical task decomposition structure and the rating scale structures along with the feedback from pilots and test evaluators. Construct validity is addressed by applying correlational and regression analysis techniques, including that of path analysis.

RESULTS AND DISCUSSION

RELIABILITY

INTERNAL CONSISTENCY

Internal consistency analyses were performed on four sets of data which included two groups of seven A-7 pilots and two groups of nine F-16 pilots. There were 167 A-7 and 221 F-16 tasks analyzed covering prelaunch, emergency, and Visual Flight Rules (VFR) functions in both tactical and non-tactical flight modes. As shown in Table 1 the internal test consistency coefficient, Cronbach's alpha (reference 4) for each data set exceeds 0.95. This measure indicates the degree to which the ratings across tasks are similar. It appears that the test items are essentially tapping a single factor or dimension; that is, EO is a unitary measure. Furthermore, with the acceptance of an internal consistency index of 0.80, the number of task elements required could be reduced by approximately one-half of the 167 items for the A-7 groups. A similar reduction in the size of the F-16 task inventory would also be permitted with this lower reliability requirement. The similarities in consistency measures for the A-7 and F-16 data sets provide evidence that the SOMA methodology has produced inventories and structures that are equivalent in terms of internal structure and factor composition.

\- -

Insert Table 1 Here

\- -

INTERRATER RELIABILITY

Analyses examining the agreement among raters were performed on operability scores for the A-7 and F-16 data. Figure 5 summarizes the interrater reliability plotted as a function of subject sample size. For the sample range of four to seven (A-7) and four to nine (F-16), subjects were randomly dropped from each data set before reliabilities were recalculated. The reliability of EO (0.60) appears acceptable at and above a sample size of six. Reliability drops more rapidly as N is reduced below six observers. This finding supports the notion that factors such as training level and perceptual set for each rating scale are sufficiently controlled if N is equal to or greater than six.

- -

Insert Figure 5 Here

- -

Figure 6 presents similar ascending reliability coefficients for the two F-16 data sets as N is increased from 4 to 9 for the C and the TE scales. On the other hand, PW reliability indices are relatively low (ranging from 0.25 to 0.58) and vary considerably between test groups. The test-retest reliability for a subset of six pilots who performed both tests was very high (0.96). That is, although individual pilots retained stable workload criteria (test-retest), there is substantial variation in criteria between pilots. If this is the case, then either more effort must be made to normalize pilot workload criteria or, as an alternative, the maximum number of pilots should be employed to extract an acceptable level of interrater reliability on pilot workload measures.

- -

Insert Figure 6 Here

- -

INTERRATER CONCORDANCE

Concordance analyses (reference 6) were performed for the A-7 and F-16. As shown in Table 2, concordance remains above 0.90. These results confirm the hypothesis that an averaged matrix can be employed to generate the operability matrix for each data set. In addition, the overall concordance analysis in which matrix ratings for F-16 and A-7 groups were combined indicates a concordance ratio (Kendall's W) of 0.96. This result indicates that the scale structure for PW, TE, and EO are common across pilot groups. Thus, a standard matrix for operability can be constructed which is suitable for all of the EO tasks involving single seat fighter and attack aircraft.

- -

Insert Table 2 Here

- -

INTRA-RATER RELIABILITY

Two groups of F-16 data were analyzed for test-retest reliability (reference 5) for 189 of the 221 items in the completed inventory (Table 3). The 32 pre-flight task items were excluded from the inventory due to computer software limitations. These items were assumed to be the most reliable of the task set. The calculated reliability estimates are thus conservative. Note that in Table 3 the six pilots' test-retest reliability for EO was 0.72.

That is, over 50% of the variance was accounted for by these tests in achieving the EO scores. With the assumption that no additional training or experience was obtained between tests, the resultant reliability coefficient would still be viewed as an acceptable test-retest level. However, all of the pilots had additional experience with the aircraft system. This added flight time probably alters pilot evaluation criteria, thus, decreasing reliability. The achieved test-retest coefficient is thus viewed as a conservative estimate.

- -

Insert Table 3 Here

- -

TEST-RETEST RELIABILITY FOR C, PW, TE

Reliability coefficients were forced for the ordinal C, PW and TE scale results over the F-16 189 item inventory. The TE and PW scale scores (Table 3) within pilots remained stable across tests (reliability coefficients of 0.87 and 0.96, respectively). The C ratings, on the other hand, were less stable (reliability coefficient = 0.63). The above findings contrast with the interrater reliability indices obtained on these scales. PW varied between pilots and remained stable within pilots while C ratings tended to be similar between pilots but varied within pilots. These contrasts provide evidence that external factors such as examiner input or biases influence scales such as C. It is also apparent that pilots tend to form a constant concept of workload, but each pilot maintains his own set of criteria for this scale dimension.

TEST-RETEST MATRICES

Operability matrices constructed by the six F-16 pilots who participated in both tests were analyzed for test-retest reliability by forcing the data with the standard reliability analysis. The obtained reliability coefficient (0.99) further strengthens the argument that a standard rating matrix is achievable for fighter and for attack single seat aircraft SOMA analyses.

VALIDITY

FACE VALIDITY

Face validity is extremely important to any evaluation tool which depends on user input. The iterative MAU process used to develop the SOMA model incorporated operational pilots into the process throughout the development of the model for each airframe. The following is a description of the hierarchical decomposition development process and the scale generating process that illustrates user integration into these processes. Mission requirements and pilot functions were task analyzed; a preliminary hierarchical decomposition was then developed. Pilots were then requested to evaluate the structure. The pilots' feedback on the fidelity and utility of the hierarchy served as the basis for any restructuring of the hierarchy. The second iteration was also submitted for pilot review. This process continued until pilots were satisfied that the decomposable hierarchy accurately reflected their mission requirements and that the structure flowed in correct mission sequence.

Operational pilots were also used in the SOMA process to generate the C, PW and TE rating scales. These scales were evolved through a process similar to the development of task hierarchies. Scales and definitions were designed to cover the range of operability along the three dimensions. The definitions were structured to reflect mission (C), system (TE) and man (PW) dimensions and at the same time maximize the orthogonality among the scales. A four interval forced choice scale was selected to sample more adequately the full range of the scale. The scales were then evaluated by pilots to ensure that the scales were useable, meaningful and reflected the total operability function. High test-retest, internal consistency, and interrater agreement on each of the scales as well as the output measure, EO, suggest that the scales and the hierarchies are understood and useable by the pilots.

CONSTRUCT VALIDITY

The statistical procedure of path analysis (reference 7) was used to evaluate the construct validity of the SOMA model. Two path analyses were calculated for the F-16 data sets. The separate tests were performed to compare the results for consistency and stability of both path coefficients. Table 4 presents the data developed from multiple regression analyses which provided the input data for the path analysis model illustrated in Figure 7. The simple partial correlations (r_{ij}), which represent the total covariance on any path, are taken when C, PW and TE are regressed on EO. The direct path coefficients (P_{ij}) are standardized betas with the lone exception being X_3 X_4 where r_{34} and P_{34} are equal using the model presented in Figure 7. The indirect causal effects are formed through combination rules dependent on the paths specified by the model. Total causal effects are simply the algebriac summation of indirect and direct effects. For instance, in X_2 X_4, total causal variance = 0.22 + (-0.12). Other factors which contribute to the simple partial correlations (r_{ij}) which are considered non-causal in terms of the specified model are derived by algebriac subtraction of the total covariation from total causal covariation. The model represented in Figure 7 was constructed by hypothesizing that PW, TE and C formed the basis for the EO dimension used in development of the EO scale (Figure 4). The C dimension is not included in the PW by TE matrix. However, a direct path from C to EO was hypothesized to obtain a complete path analysis model. The primary input from C to EO is assumed to occur through PW and TE. TE is assumed to be causally linked to PW because the equipment in the system is fixed and should form one anchor in the development of the PW rating.

- -

Insert Table 4 Here

- -

- -

Insert Figure 7 Here

- -

Findings which confirm the validity of the hypothesized MAU model would meet the following four requirements:

1. EO is entirely determined by the three input dimensions
2. Most, if not all, of the PW, C and TE variability is externally determined

3. C demonstrates no direct causality with EO but demonstrates substantial indirect causal covariance

4. PW and TE are independent.

The results of the path analysis appear to satisfy the requirements for the hypothesized MAU. For requirement 1 most of the variance in effective operability is internally determined (Group 1, $R^2 = 0.80$; Group 2, $R^2 = 0.94$). For requirement 2 the variation from external sources on PW and TE range from a low of 0.56 for PW in group 2 to a high of 0.89 for TE in Group 1. For requirement 3, C shows no direct relation with EO (Direct Causal $P_{14} = 0.03$ and -0.04) while the indirect effects of C on EO are -0.22 and 0.28 for Groups 1 and 2, respectively. For requirement 4 the relatively strong relation (0.32 and 0.68) between PW and TE, suggests that these scales need to be improved if the orthogonality assumption is to be strictly satisfied for applying the conjoint measurement approach.

One area where the PW scale might be improved is by making its scale units match those used for TE. There is a negative relation between C and TE (-0.2 and -0.35) while the C and PW relationship tends to be positive (0.22 and 0.06). This inverse relation between C and TE and between C and PW tends to explain why C relation has been difficult to incorporate previously into the model. Inspection of the PW and TE scales reveals that TE is a positive scale while PW is a negative scale. If both were positive the covariation of PW and TE with C would be in the same direction.

SUMMARY AND CONCLUSIONS

Reliability and validity analyses were performed on A-7 and F-16 data subjected to the SOMA technique. The following conclusions were reached:

- Task inventories for SOMA could be reduced by 1/2 without excessively decreasing internal consistency reliability.
- Effective operability is essentially a single dimension or single factor measure.
- SOMA methodology has produced inventories and structures that are equivalent in terms of internal structure and factor composition.
- Interrater reliability for EO, C and TE are acceptable when N = 6 or greater.
- Average rater matrices may be used to develop the EO scale.
- A standard matrix for EO can be constructed which is suitable for single seat fighter or attack aircraft.
- The EO test-retest ratio of 0.72 (accounting for over 50% of total variance) is a conservative reliability estimate.
- Workload estimates are consistent within pilots but variable between pilots.
- Reliability estimates verify that the method of construction of the task hierarchy and scales demonstrate face validity.
- The SOMA model is essentially valid.
- The PW scale should be adjusted through definitions to
  a. improve orthogonality with the TE dimension
  b. scale along a monotonic dimension

REFERENCES

1. Helm, W.R. and Donnell, J.L., "Mission Operability Assessment Technique: A System Evaluation Methodology", Technical Publication TP-79-31, Pacific Missile Test Center, Point Mugu, CA., 10 October 1979.

2. Edwards, W., "How to Use Multi-Attribute Utility Measurement for Social Decision Making", Technical Report 001597-1-T, Social Sciences Research Institute, University of Southern California, Los Angeles, CA. 1976.

3. McClelland, G.H., "Equal vs Differential Weighting for Multiattribute Decisions: There Are No Free Lunches", Technical Report, University Of Colorado Institute of Behavioral Research, 1978.

4. Cronbach, L.J., "Coefficient Alpha and the Internal Structure of Tests", Psychometrika, 16, 297-334, 1951.

5. Guilford, J.P. Psychometric Methods, McGraw-Hill Book Company, Inc., New York, 1954.

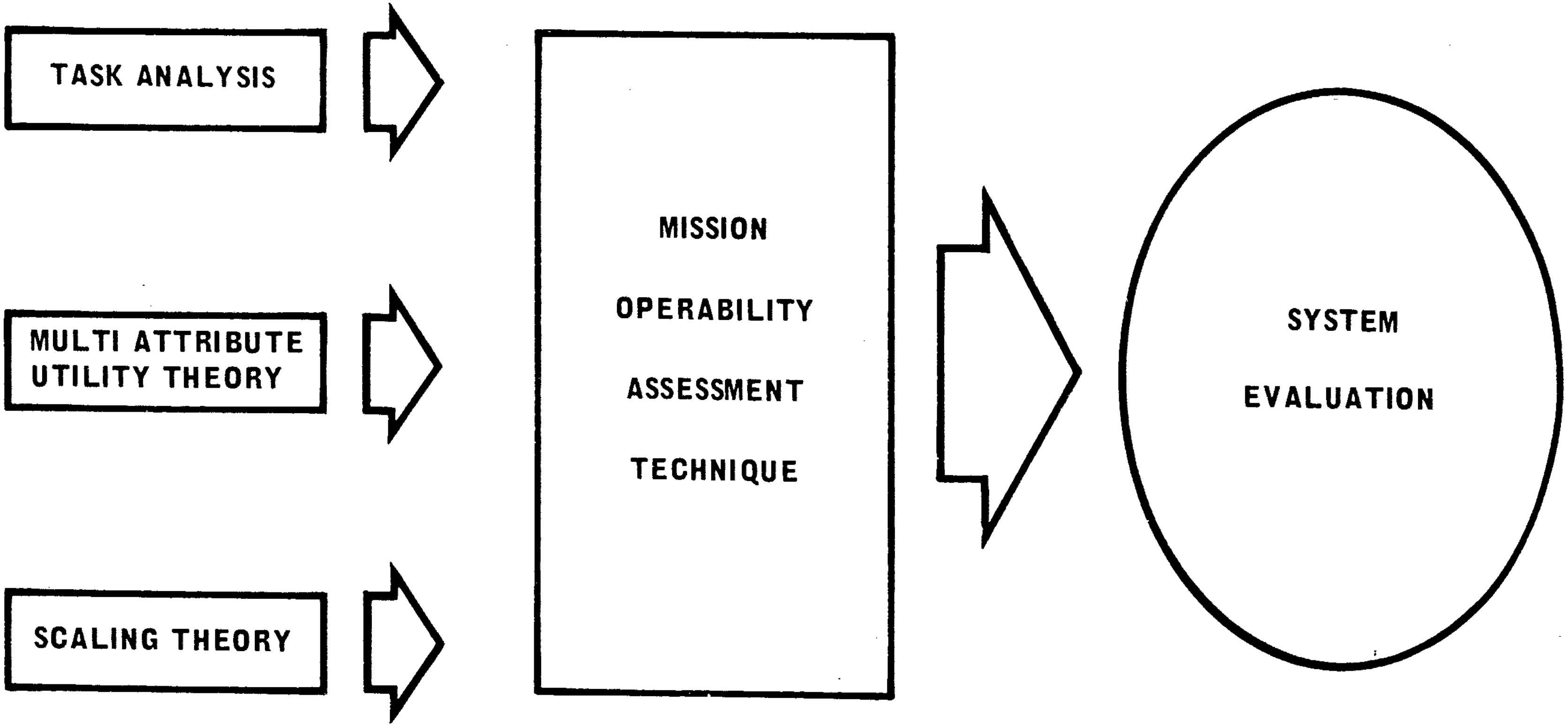
6. Hollander, M. and Wolfe, D.A. Nonparametric Statistical Methods, John Wiley and Sons, New York, 1974.

7. Duncan, O.D. Path Analysis: "Sociological Examples", American Journal of Sociology, 72, pp 1-16, 1966.

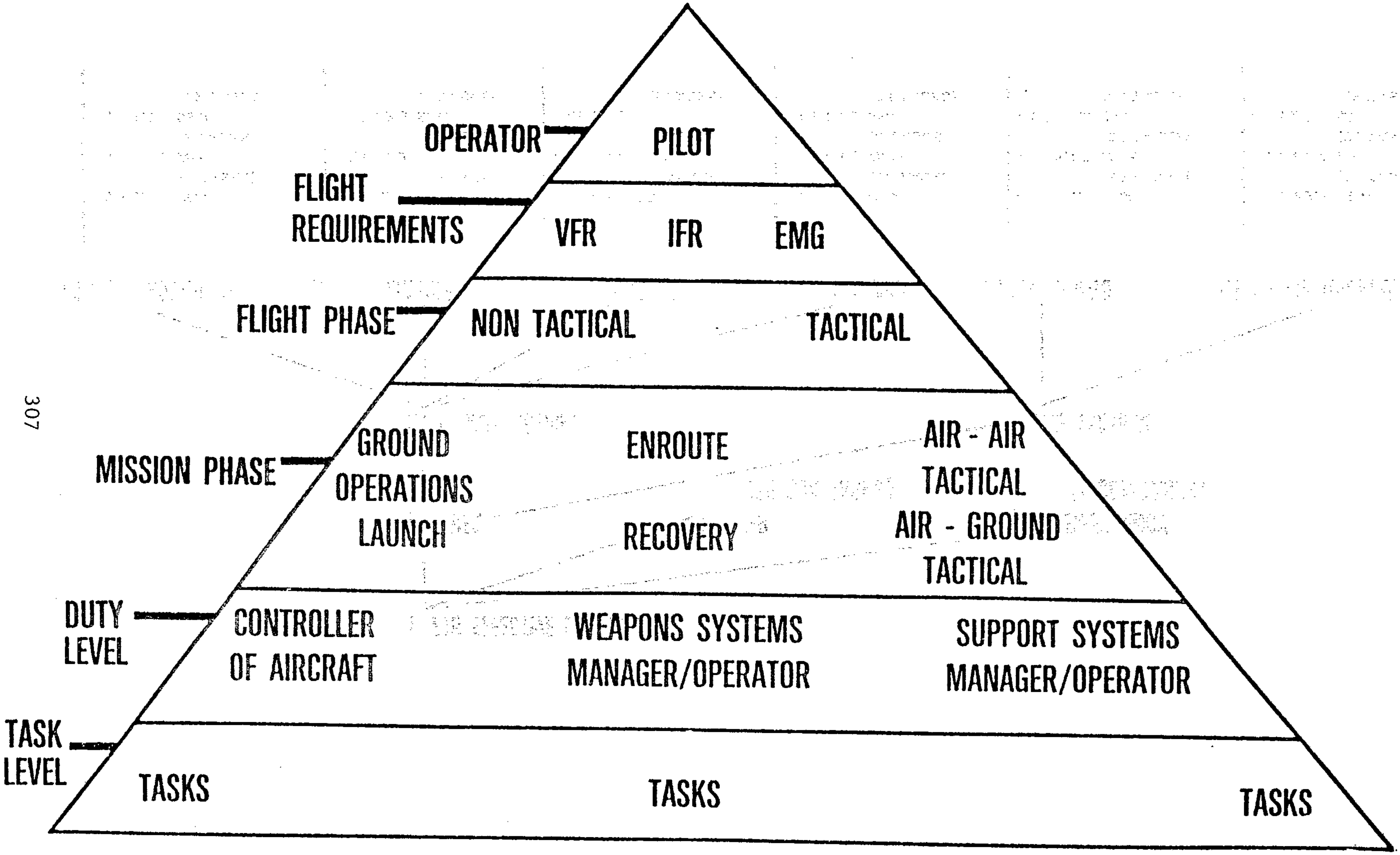
FIGURE CAPTIONS

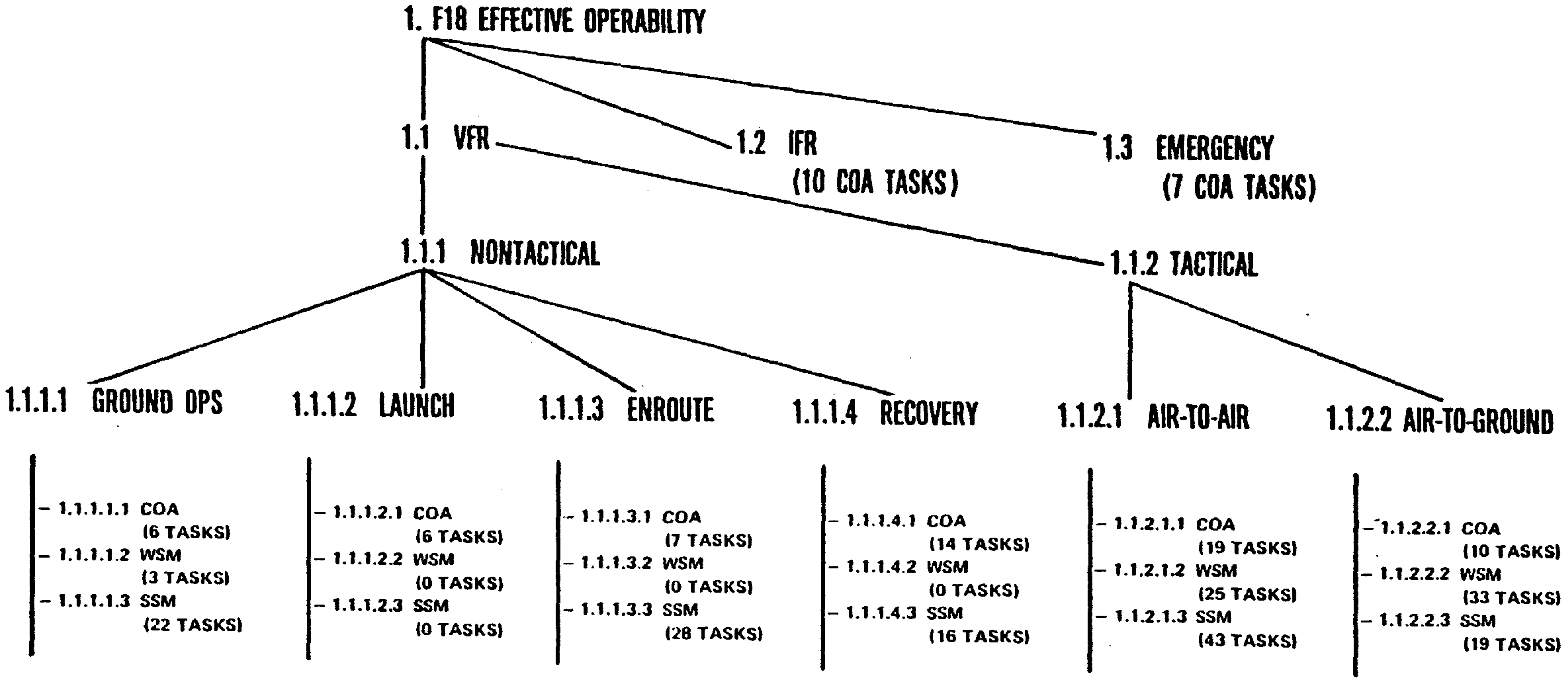
1. SOMA Developmental Model
2. Mission Task Hierarchy
3. Mission Needs Hierarchy Decomposition Structure
4. Conjoint Measurement Methodology
5. Effective Operability Reliability by Number of Raters for A-7 and F-16
6. Criticality, Pilot Workload, and Technical Effectiveness Reliability by Number of Raters for A-7
7. SOMA Path Model


????
PROB
MANCE
ER
LEMS
FOR
P


TASK ANALYSIS
MULTI ATTRIBUTE UTILITY THEORY
SCALING THEORY
MISSION OPERABILITY ASSESSMENT TECHNIQUE
SYSTEM EVALUATION

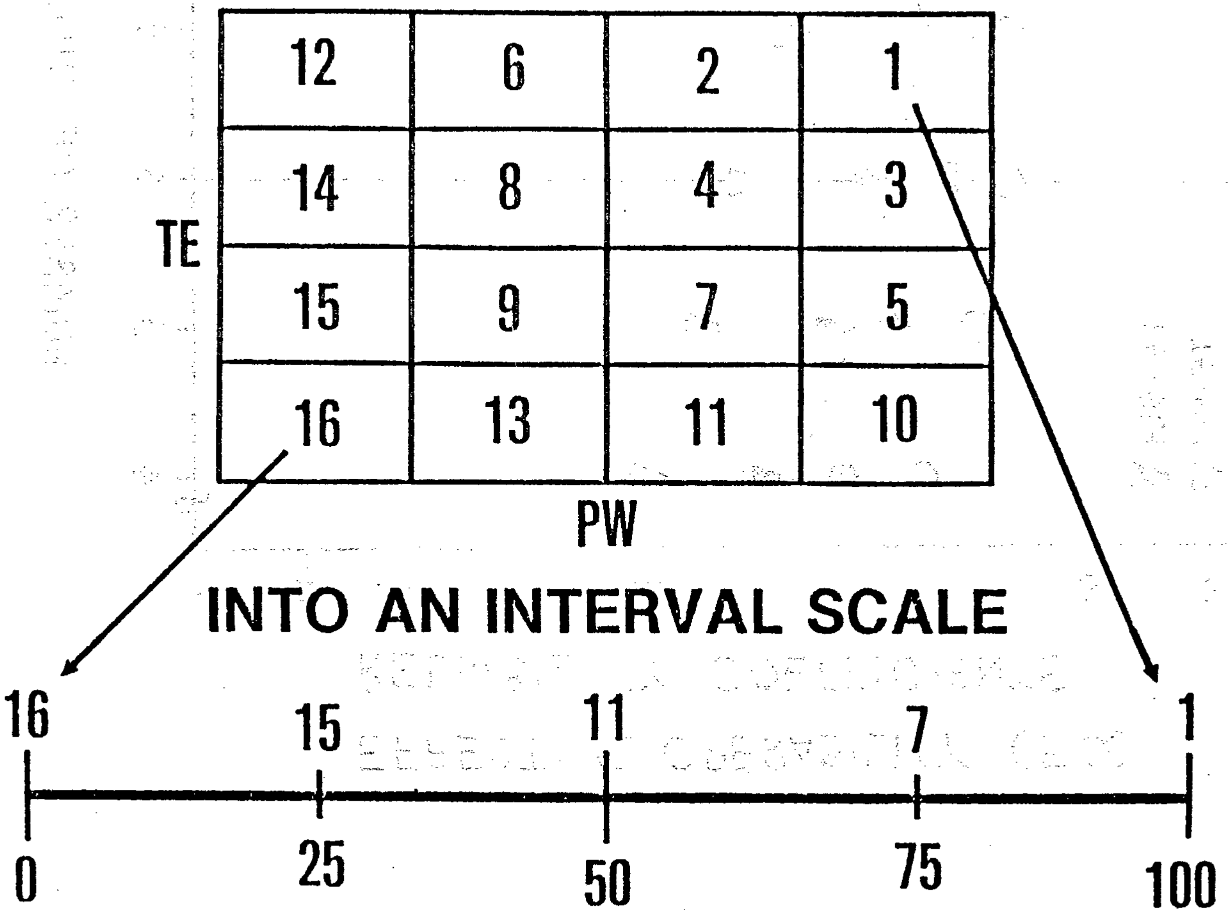
MISSION TASK HIERARCHY




1. F18 EFFECTIVE OPERABILITY
1.1 VFR
1.2 IFR
(10 COA TASKS)
1.3 EMERGENCY
(7 COA TASKS)
1.1.1 NONTACTICAL
1.1.2 TACTICAL
1.1.1.1 GROUND OPS
1.1.1.2 LAUNCH
1.1.1.3 ENROUTE
1.1.1.4 RECOVERY
1.1.2.1 AIR-TO-AIR
1.1.2.2 AIR-TO-GROUND
1.1.1.1.1 COA
(6 TASKS)
1.1.1.1.2 WSM
(3 TASKS)
1.1.1.1.3 SSM
(22 TASKS)
1.1.1.2.1 COA
(6 TASKS)
1.1.1.2.2 WSM
(0 TASKS)
1.1.1.2.3 SSM
(0 TASKS)
1.1.1.3.1 COA
(7 TASKS)
1.1.1.3.2 WSM
(0 TASKS)
1.1.1.3.3 SSM
(28 TASKS)
1.1.1.4.1 COA
(14 TASKS)
1.1.1.4.2 WSM
(0 TASKS)
1.1.1.4.3 SSM
(16 TASKS)
1.1.2.1.1 COA
(19 TASKS)
1.1.2.1.2 WSM
(25 TASKS)
1.1.2.1.3 SSM
(43 TASKS)
1.1.2.2.1 COA
(10 TASKS)
1.1.2.2.2 WSM
(33 TASKS)
1.1.2.2.3 SSM
(19 TASKS)

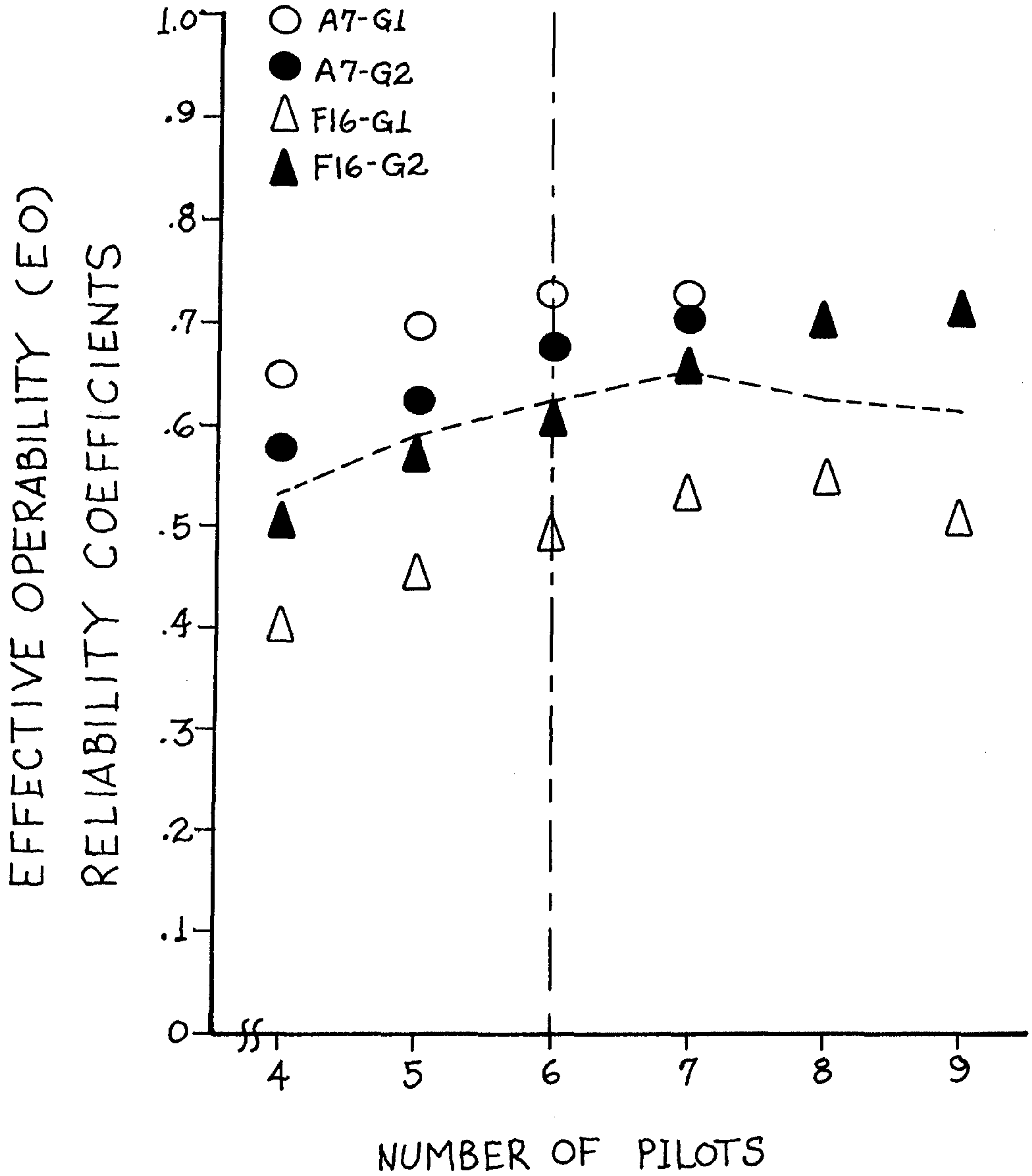
CONJOINT MEASUREMENT METHODOLOGY

COMBINES PAIRS OF PILOT RANKS ON PW AND TE

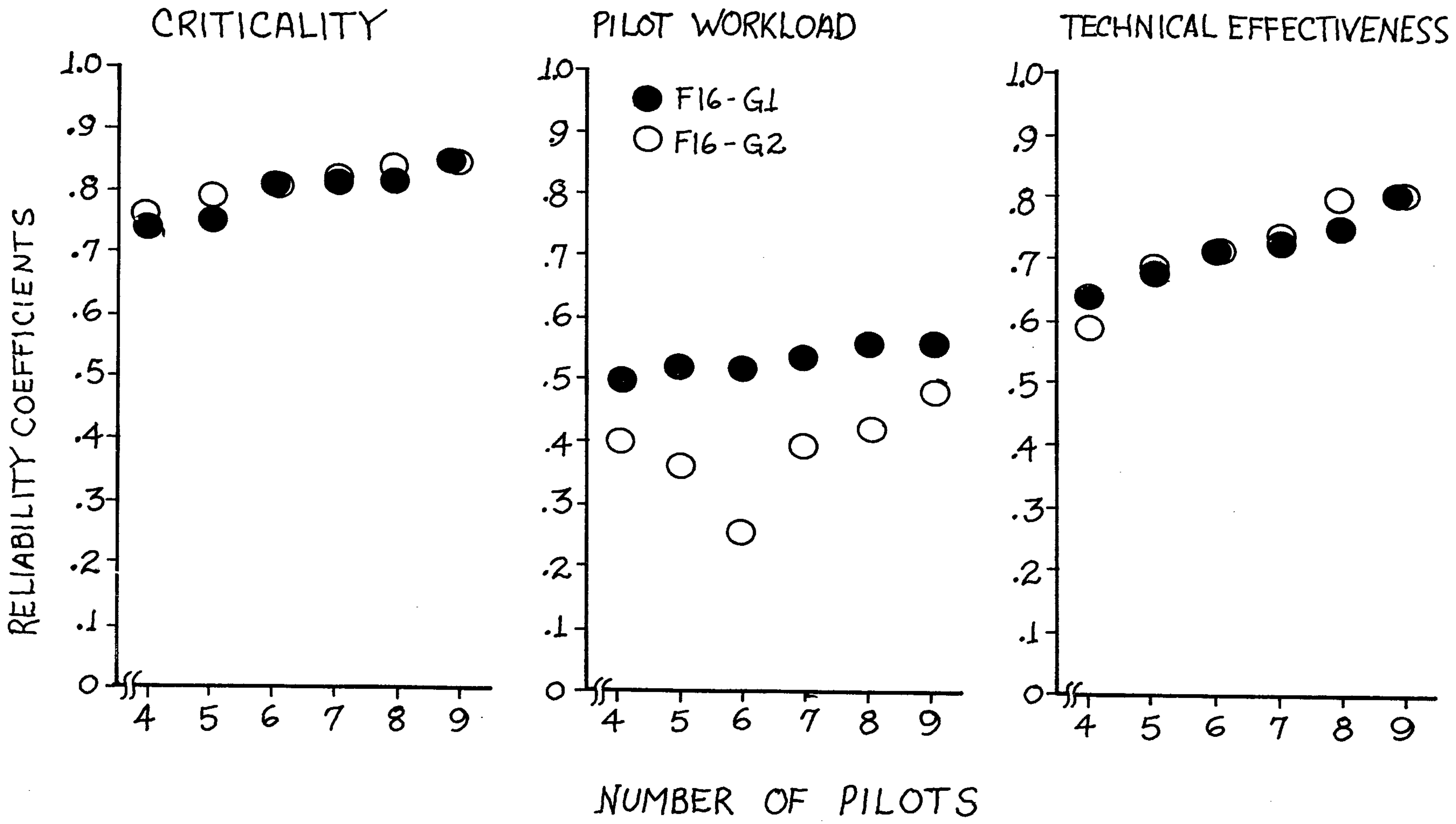


12	6	2	1
14	8	4	3
15	9	7	5
16	13	11	10

INTO AN INTERVAL SCALE



○ A7-G1
● A7-G2
△ F16-G1
▲ F16-G2
EFFECTIVE OPERABILITY (EO)
RELIABILITY COEFFICIENTS
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
4
5
6
7
8
9
NUMBER OF PILOTS


CRITICALITY
PILOT WORKLOAD
TECHNICAL EFFECTIVENESS
● F16-G1
○ F16-G2
RELIABILITY COEFFICIENTS
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
4
5
6
7
8
9
NUMBER OF PILOTS

SOMA PATH MODEL

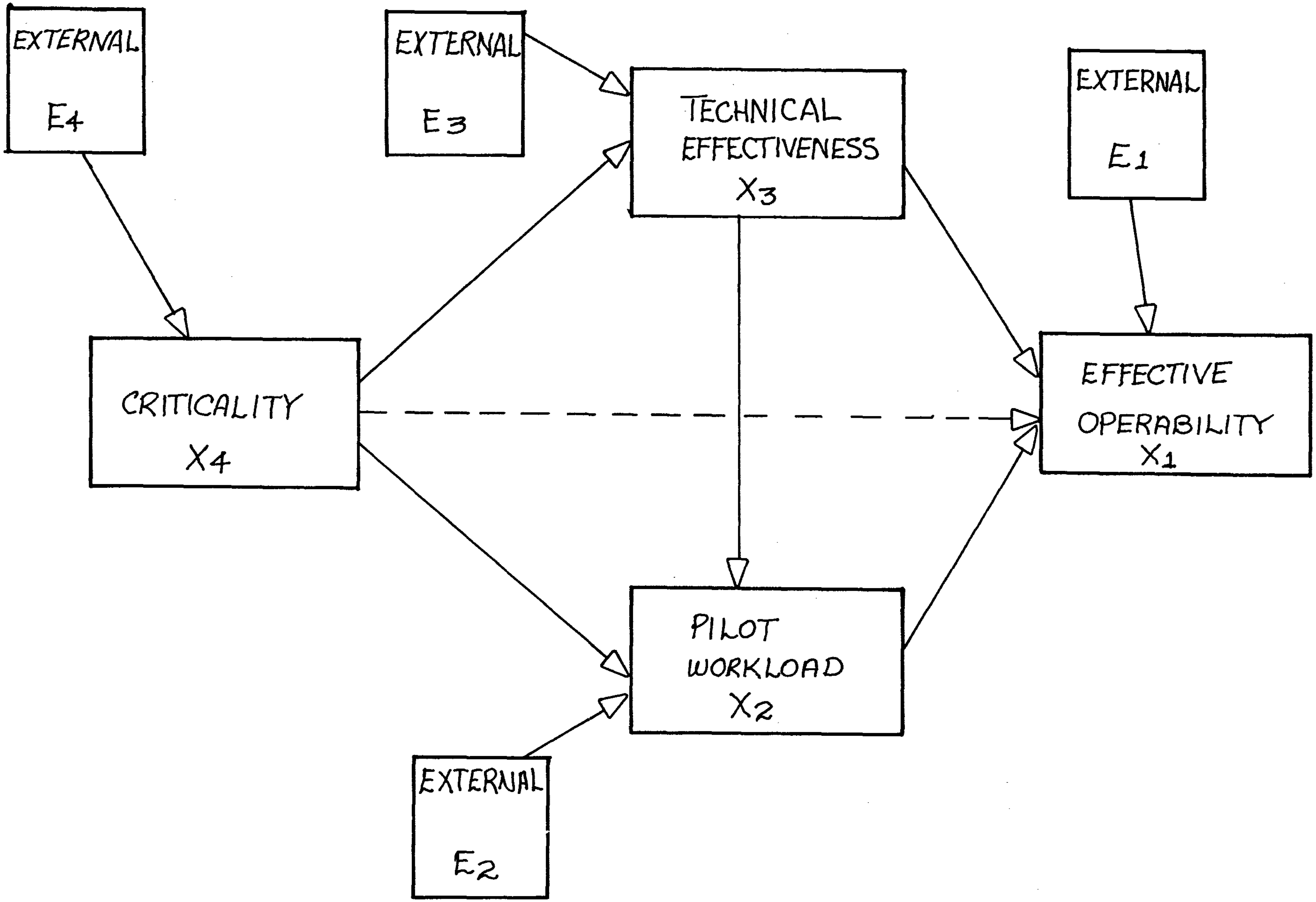


TABLE CAPTIONS

TABLE 1. SUMMARY OF INTERNAL CONSISTENCY RELIABILITY FOR A-7 AND F-16

GROUP	EO	TE	PW	C	ITEMS	PILOTS
A7-I	.96	*	*	*	166	7
A7-II	.98	*	*	*	166	7
F16-I	.98	.99	.99	.99	189	9
F16-II	.99	.99	.99	.98	189	9

*PW, C AND TE PILOTS' RAW SCORES WERE NOT AVAILABLE FOR THE A-7 ANALYSES

TABLE 2

OPERABILITY MATRIX INTER-RATER CONCORDANCE RATIOS (KENDALL'S W)

GROUP	N	KENDALL'S COEFFICIENT OF CONCORDANCE W
F-18	6	.94*
A7-79	12	.92*
A7-G1	7	.96
A7-G2	7	.98
F16-G1	9	.98
F16-G1	9	.99

*FROM HELM AND DONNELL, 1979

TABLE 3.
F-16 TEST-RETEST RELIABILITY SUMMARY

				GROUP 1		GROUP 2		
	TEST	N	ITEMS	X	S.D.	X	S.D.	RELIABILITY COEFFICIENT
1.	EFFECTIVE OPERABILITY	6	189	59.37	12.46	64.18	16.92	.71
2.	CRITICALITY	6	189	4.39**	1.30***	4.25**	1.50***	.63*
3.	TECHNICAL EFFECTIVENESS	6	189	2.27**	1.08***	2.44**	1.48***	.87*
4.	PILOT WORKLOAD	6	189	2.52**	1.07***	2.47**	1.29***	.96*
5.	OPERABILITY MATRICES	6	16	-	-	-	-	.99*

*ORDINAL DATA
**MEDIANS
***SEMI-INTERQUARTILE RANGE

TABLE 4. PATH ANALYSIS SUMMARY

	BIVARIATE RELATIONSHIP	TOTAL COVARIANCE	CAUSAL			NON CAUSAL	EXTERNAL SOURCE VARIATION	
			DIRECT	INDIRECT	TOTAL			
F-16 GROUP 1	X_3X_4	-.318	-.318	-	.318	-	1 .20	2 .87
	X_2X_4	.106	.224	-.118	.106	-		
	X_1X_4	-.247	-.031	-.216	-.247		3 .90	4 1.00
	X_2X_3	.302	.373	-	.373	-.071		
	X_1X_3	.858	.766	.101	.867	-.009		
	X_1X_2	.499	.271	-	.271	.228		
F-16 GROUP 2	X_3X_4	-.345	-.345	-	-.345	-		
	X_2X_4	-.174	.063	-.273	-.174	-	1 .01	2 .55
	X_1X_4	-.311	-.035	-.276	-.311	-		
	X_2X_3	.664	.686	-	.686	-.022	3 .88	4 1.00
	X_1X_3	.892	.540	.351	.891	.001		
	X_1X_2	.876	.511	-	.511	.365		

A SELF-CONTAINED, MAN-BORNE BIOMEDICAL INSTRUMENTATION SYSTEM IN THE FLIGHT TESTING OF NAVAL WEAPONS SYSTEMS

CDR Douglas W. Call, MSC, USN *
LCDR David M. Kelly, USNR-R +
Douglas G. Robertson ++

Flight Test Division
Pacific Missile Test Center
Point Mugu, California

ABSTRACT

The historical lack of aircrew physiological data has prevented its use in the definition and evaluation of airborne man-machine systems. Recognizing this deficiency, a joint Navy and Air Force program has resulted in the development of a self-contained man-borne In-Flight Physiological Data Acquisition System (IFPDAS). This system is capable of obtaining and recording up to 32 channels of physiological and engineering parameters for flight durations to four hours. A ground based microprocessor provides computer analysis of this data. At present, the Navy is establishing a methodology for the employment of the system, developing analytical software, and collecting operational flight data. Further plans for the IFPDAS include collecting and processing multi-parameter physiological and environmental data from aircrewmen to more thoroughly describe the tactical flight environment and thus support proper development of aircrew life support and weapons systems.

THE PROBLEM

Currently operational aircraft such as the F-14 are capable of performance which in many flight regimes exceeds the pilot's ability to sustain normal physiological function. The F/A-18 undergoing test and evaluation at the Pacific Missile Test Center (PACMISTESTCEN) presents a still greater operator challenge. Even with sophisticated life support systems, man is often the limiting factor in the optimal employment of an aircraft performance capabilities. From a weapons system standpoint, the aircrewman's psychological abilities also present a significant constraint. Man's limited sensor modality, signal processing capability, and pattern recognition techniques are no match for today's accelerating technology.

While considerable research, analysis, and testing is conducted as part of aircraft and weapon systems development to determine the equipment's response to the environmental and operational demands, relatively little is done to characterize the operator's psycho-physiological responses. The design of a crew station in an airborne system is typically an extrapolation from an existing system tempered with subjective evaluations from a small sample of system operators. With existing techniques it is not currently possible to determine in an objective manner, to what extent the "human factor" limits overall system performance.

This inability to measure the operator's response to the tactical flight environment presents an additional problem to the system designer. The standards and specifications which control the development of life support, controls and display systems provide only limited guidance with respect to operator workload and boundaries for the operator's physiological response. This situation hampers the optimization of the man-machine interface and, thereby, limits the performance of future airborne systems.

History

A considerable amount of in-flight monitoring of physiological parameters was conducted by NASA throughout the 1960's. Most of this data was collected during special test flights and not under typical tactical conditions. In 1965, however, NASA researchers under two separate projects instrumented Navy pilots involved in flight operations in Vietnam[1,2] While the aggregate of these two projects was less than fifty flights the results unexpectedly indicated that from a physiological standpoint, carrier launch and recovery operations were more disruptive to the pilots than were combat operations. Observations such as these and the requirement for improved human engineering design standards and specifications prompted an increasing interest in in-flight physiological monitoring.

A U.S. Navy developed man-mounted physiology data recorder, the Bio-Pack was used in the late 1960's. This unit, as described by Horrigan et.al[3] employed a small 8 channel reel-to-reel tape recorder to store up to 40 minutes of multiplexed data. Parameters obtained were electrocardiograph (ECG), body temperature and voice from the aircrewman plus cockpit accelerations and temperatures. The Bio-Pack recorder with its battery power source was packaged in a 6" x 2" x 10" container and weighed 4.7 pounds. It could be placed conveniently on the pilots knee board or carried in the map case or any adjacent cockpit area. Over 30 successful Bio-Pack instrumented flights were conducted. Data from these tests exhibited close correlations between the stressful events of the flights and the aircrew's physiological responses. The investigators were pleased with the Bio-Pack's performance but reported plans to extend the 40 minute recording capability and to develop a smaller lighter package. Such a unit would more closely meet the design objectives of non-interference with the aircrewman, compactness, reliability, and versatility. The U.S. Air Force at Brooks Air Force Base attempted to meet these criteria in 1975 with a physiologic recorder purchased on contract from SCI Systems Inc. of

---

* Head, Crew Systems Branch
\+ Naval Reserve Aviator
++ Physical Scientist, Crew Systems Branch

Houston, Texas. This recorder, termed IFPDAS, provided for continuous simultaneous recording on cassette tape of seven channels of physiological and environmental data. These include ECG, expiratory flow, pO_2 (inhaled/exhaled O_2 concentration), G_z, acceleration (vertical to the pilot), cabin pressure, voice information and a time code. The system as described by Morgan[4] was intended to determine breathing system design parameters and investigate other biomedical requirements of flight. Navy aircrewmen at the PACMISTESTCEN were monitored with the IFPDAS unit while flying A-4, F-4 and F-14 aircraft. This data along with measurements obtained from Air Force C-130 & T-38 pilots were reported in 1975, 1976 and 1977.[4,5,6] During these flights, the components of the IFPDAS were consolidated into pockets on the pilot's survival vest. This eliminated the complicated umbilical connections required to accomodate possible ejection and man-seat separation contingencies. Thus air-crewmen began to more readily accept the slight encumbrance produced by the recorder components and became more willing to wear the IFPDAS on any mission.

One of the most valuable applications of IFPDAS was to measure breathing requirements from pilots of several different tactical aircraft under various flight conditions to determine if current oxygen delivery standards were realistic. One study[6] noted that expiratory flow rates from pilots in T-38, A-4, F-4 & F-14 aircraft typically ranged from 12.5-23.1 liter/min. on take off and as high as 28.2 liters/min. while landing. Of major significance was the finding that measured increases in minute ventilation often exceeded levels commonly referenced in breathing system design parameters[7]. Thus there are times during stressful flight events when the current USAF O_2 delivery standard of 25.1 liters/min and the USN standard of 28.0 liters/min are exceeded.[7] These standard values were usually obtained from altitude chamber situations. Morgan et.al[5] recommended that future standards require oxygen systems to deliver up to 60 liters/min during transient periods of extreme need, in addition to the routinely encountered levels of 20-25 liters/min during normal flight. Morgan further suggested that when new missions or aircraft were considered in-flight pulmonary monitoring should be conducted to assure that the respiratory life support systems be truly able to support the aircrew.

In 1976, the USAFSAM repackaged and integrated the IFPDAS for flight in Naval aircraft. The new recorder, termed IFPDAS II, was flown on approximately 30 flights at the PACMISTESTCEN between 1976 and 1979 in T-2, TA-4, A-7, F-4 and F-14 aircraft in a wide variety of flight missions including missile runs, strafing and weapon release. Functional capability of the IFPDAS was demonstrated on flights to airspeeds of mach 1.5 and during maneuvers producing up to +5 G_z. While these tests were underway the USAF and the USN made plans to develop an entirely new IFPDAS with even greater capabilities.

In 1977, the PACMISTESTCEN Microelectronics Laboratory and Weapons Instrumentation Division began work on the third generation IFPDAS. Previous programs had shown that the collection of high quality data was only one part of the physiological assessment process. Manual techniques for tabulation and analysis of these large quantaties of raw data were unacceptable. In addition to the sensors and recorder package common to all previous systems, the goal of the IFPDAS III development program was to provide a data analysis capability. The heart of this system was to be a small mini computer which could employ software controlled pattern recognition and statistical analysis routines to large volumes of flight data.

IFPDAS III was completed in 1979. It then entered an extensive test and evaluation process directed at validation of the design and environmental qualification (calibration of all the system components). Preliminary efforts were also made to develop basic software for the system.

IFPDAS III

IFPDAS III components are identified in a functional flow diagram (Figure 1). The man mounted self-contained recorder weighs just 3 1/2 lbs. and its small size (6" x 5" x 2") allows it to fit comfortably in the aircrewman's survival vest. Additional man-carried components of IFPDAS III include a pulmonary flow unit and a cardiothermal module.

The airborne recorder is capable of monitoring, conditioning and recording up to 20 channels of data for up to four hours; or, when combined with the cardio-thermal module, it can record a mix of 12 channels analog and 32 channels digital data. The data are multiplexed on a four track magnetic tape cassette. The specific data currently available on the Navy IFPDAS III include expired respiratory gas flow rates, ECG/heart rate, skin temperatures, acceleration, cabin pressure, voice annotation and a timing signal.

ECG/Heart Rate: Both ECG and a digital record of heart rate are recorded from a 3 lead chest mounted electrode harness.

Expiratory Flow Rate: A sensor in the expiration port of the Navy A-13A oxygen mask sends a pneumatic signal to a flow transducer in the pulmonary flow module which in turn produces an analog electrical output.

Skin/Cockpit Temperatures: Up to 8 skin and 2 cockpit thermistors provide inputs to the cardiothermal module. Output to the recorder is a digital signal.

Voice: All communications are monitored and recorded by means of a sidetone tap into the headset side of the mask interphone cord.

±Gz Accelerations: ±Gz accelerometer (-3 to +10 G_z range) is located within the recorder package.

Cockpit Pressure: An aneroid barometer in the recorder module measures ambient cockpit pressure (0-760 mm Hg).

Timing: An internal clock generates a 16 BIT timing code to correlate all measured parameters.

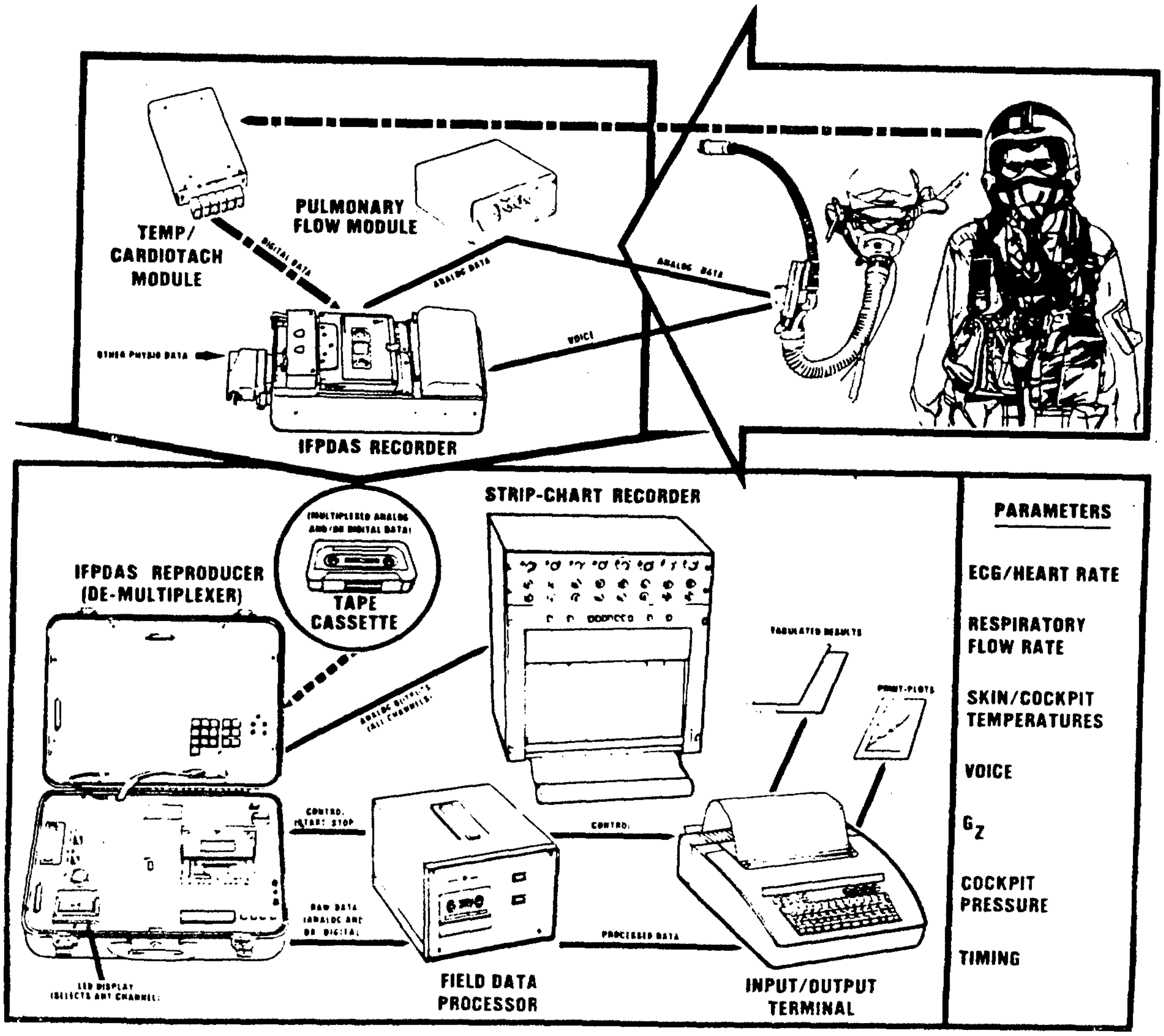
The recorder can monitor, condition and record data for 4 continuous hours. The conditioned signals from the sensors are time multiplexed by three 8 channel analog multiplexers and recorded on a standard 4 track magnetic tape cassette using a pulse duration modulation (PDM) format. The 4th track is used for sidetone voice recording.

The ground based components of the system include a data reproducer playback device, a portable field data processor, teleprinter and laboratory data processor. For data analysis the IFPDAS reproducer converts the PDM signals from the recorder to analog signals. The voice channel is fed into an audio amplifier and speaker within the unit. Using the reproducer's control options, any of the analog signals can be:

a. Displayed as a voltage on a liquid crystal display in the reproducer.
b. Fed into a strip-chart recorder
c. Output to the field data processor.

The reproducer is capable of being operated in a pre-programmed time mode in which the start and stop times of the required portion of the data may be entered via the keyboard and an automatic tape search is initiated and controlled by the processor.

In the field data procesor, analog data from the reproducer is converted to digital data by an A to D converter. A control tape in the processor controls the output of the reproducer; i.e., rewind, fast forward, play. The microprocessor in the field data processor was designed with integral analog multiplexers (32 channels), a 10 bit analog-to-digital converter, and digital input/output ports. The device uses 11 kilobytes of erasable-programmable read-only memory and 24 kilobytes of random-access memory. The field data processor allows manipulation of the data using test oriented BASIC language. Processed data is displayed on a teleprinter. The data can be printed in either tabular or plot formats. The IFPDAS recorder does not require aircraft modification or installation; no aircraft resources are required, no telemetry channels or aircraft operational priorities are necessary and the system is not aircraft-type specific. Therefore, all support features common to usual test installations are eliminated.



USN IFPDAS FUNCTIONAL FLOW DIAGRAM

SYSTEM APPLICATION

Concurrent with the development/validation process, the IFPDAS III unit has been used to collect and analyze data for several ongoing U.S. Air Force and Navy programs. Following is a brief description of these projects and the role of the IFPDAS.

USAF Cockpit/cabin temperature data, aircrew heart rates in C-130, C-5A, C-141, F-15 aircraft. Environmental control systems evaluation, aircrew heart rate and skin temperatures, A-7K aircraft.

Navy Crew station accelerations and crew heart rate responses on SEPTAR Target boats. General aircrew workload in A-7, A-6 and F-14 aircraft.

Current Plan During the ongoing engineering development phase of this program, IFPDAS capabilities will be expanded in three specific areas. These include:

a. Gaining experience in collection of laboratory and in-flight data to verify and refine data acquisition methodology;

b. Development and verification of software routines to allow more sophisticated data analysis;

c. Gaining operational experience by deploying the IFPDAS to a remote operating location (deployed unit) or aboard an aircraft carrier for extensive use away from readily available maintenance facilities.

This phase will conclude with the formal documentation of the procedural and analytical methodology for the effective employment of this testing tool. Then the IFPDAS can be used to help answer such vital questions as:

1. How is the aircrewman's efficiency affected by present heavy, bulky, personal protective equipment and are life support systems really supportive?

2. Do new weapons systems platforms exceed the physiological capabilities of the aircrewmen?

3. Can information on operator workload be used to develop more efficient training procedures and simulators?

ACKNOWLEDGMENT

This project was supported by the Naval Medical Research and Development Command, under Research Work Unit MF 58524024-0005. The views presented in this paper are those of the authors. No endorsement by the Department of the Navy has been given or should be inferred.

References

1. Lewis, C.E., Jones, W.L., Austin, F. and Roman, J. Flight Research Program: IX. Medical Monitoring of Carrier Pilots in Combat -- II. Aerospace Med. 38:581-592, 1967

2. Roman, J., Older, H. and Jones, W.L. Flight Research Program: VII. Medical Monitoring of Navy Carrier Pilots in Combat. Aerospace Med. 38:133-139, 1967.

3. Horrigan, D.J. Jr., Martin, J.D. and Engbrecht, W.J. A Self-Contained System for In-Flight Acquisition of Physiological Data. Naval Missile Test Center TP-70-61(U), 29 October 1970.

4. Morgan, T.R. Development of an In-Flight Physiological Data Acquisition System. Proceedings of SAFE Symposium 21-25 September 1975, San Antonio, Texas. SAFE Association, Canoga Park, California.

5. Morgan, T.R., Baumgardner, F.W., Crigler, J.C., Reid D.H. and Tays, M.A. Preliminary Analysis of Available In-Flight Respiratory Data, Report SAM-TR-77-20. USAF School of Aerospace Medicine, Brooks AFB, Texas. December 1977.

6. Morgan, T.R., Reid, D.H. and Baumgardner, F.W. Pulmonary Ventilation Requirements Evident in the Operation of Representative High Performance Aircraft. Preprints from the 47th Annual Scientific Meeting of the Aerospace Medical Association pp. 156-157, Bal Harbour, Florida. 10-13 May 1976.

7. MILITARY SPECIFICATION MIL-D-19326F of 18 October 1978 Design and Installation of Liquid Oxygen Systems in Aircraft, General Specification for

8. Call, D.W., Robertson, D.G. and Kelly, D.M. In-Flight Physiological Monitoring of Tactical Jet Aircrews. Proceedings of 18th Annual SAFE Symposium. 12-16 October 1980, San Diego, California. SAFE Association, Canoga Park, California

Physiological and Performance Parameters as
Indices of Pilot Workload
An Analysis of Data from the AFTI F-16 Project

by

Wg. Cdr. D. C. Reader, RAF
USAF School of Aerospace Medicine
Brooks AFB, Texas 78235

INTRODUCTION

One important new project in fighter aircraft technology is the Advanced Fighter Technology Integration (AFTI) F-16. The performance of this airplane will be enhanced by the installation of aerodynamic canards to the underside of the engine air intake. Deflection of these surfaces together with the rudder, elevators, and flaps will allow the aircraft to translate vertically and laterally or hold constant pitch or yaw attitudes. This will reduce the vulnerability of the aircraft without affecting its firepower. These computer-controlled aerodynamic surfaces will greatly increase the performance of the airplane. Thus the limitation to the weapon system may well be the pilot; by his inability to sustain high accelerations in the vertical and lateral axes and his mental processing time which could be considerably longer than that of the aircraft computers.

In order to reduce, as much as possible, this limitation to the overall weapon system, the USAF School of Aerospace Medicine was asked to devise methods to measure the workload of the pilot to predict when limitations would occur and to advise what improvements were necessary to enhance the performance of the pilots.

Initially, much of the development work would be performed in simulators; a fixed base simulator (FBS) at General Dynamics, Inc., at Fort Worth, Texas, and the Dynamic Environmental Simulator (DES) at the Air Force Aerospace Medical Research Laboratory at Wright-Patterson AFB, Dayton, Ohio. During this simulation effort, physiological and performance data were recorded from the pilots and analyzed at the USAF School of Aerospace Medicine. This paper describes the data, the analyses and the conclusions drawn.

METHODS

The physiological data recorded were the electrocardiogram (EKG) and electromyograms (EMG) from the right arm and left neck areas. Raw tracking error data from the simulator computers were also recorded as direct measures of performance. These data were amplified, filtered and recorded on an FM multi-channel tape recorder.

These data were then digitized, summarized and subjected to analysis of variance. Throughout the simulations hand written logs in IRIG B Code were made of pilot activity and the tasks performed by each. The logs were then used to compare the data throughout the various sequences of activity.

From the EKG data, heart rate and rate variance were calculated and the amplitude of the R wave of the EKG complex was measured and its variance calculated. The EMG data were analyzed for center spectrum frequency and total voltage. The tracking data were analyzed for amplitude and variance.

RESULTS

In the fixed base simulator experiments the pilots flew a variety of manual control tasks, tracking air-to-air and air-to-ground targets. Various aspects of the AFTI software were introduced one by one in semi-random order and their effects noted.

Heart rate - Figure 1 graphs the mean heart rates of the subjects for four tracking tasks (air-to-air and ground) as speed was varied. The trends show that the heart rates rose as the speed increased. In this situation, the physical workload remained almost constant but decisions and tracking effort became more demanding. Such an increase in mental workload was shown by the small rise in heart rate.

Heart rate variance - Figure 2 plots the heart rate variance (root mean square, RMS) of the subjects grouped with speed increase. As with heart rate, an overall rise was seen. It should be noted that the downward trend of task 3 & 5 above a simulated speed of 350 kts was not statistically significant.

R wave amplitude - The amplitude of the R wave for the subjects grouped is shown in Figure 3. As speed increases, the R wave declines overall. Admittedly, tasks 6 & 7 showed a small rise over a narrow speed range. The trend for tasks 8 & 9 failed to show any significant difference between 300 and 500 kts simulated airspeed.

R wave amplitude variance - Figure 4 shows how R wave variance altered with speed increase. Overall, an increase in variance was observed.

EMG - The arm electromyogram recordings showed considerable variability. With three of the tracking tasks, the center frequency fell as speed increased; with the fourth no change was observed. The data recorded from the neck site were corrupted with high frequency interference from switch gear in the simulator and could not be used.

Tracking data - The variance of the tracking data also showed considerable variability. In the first three tracking tasks it increased; in the last it showed little change.

The Dynamic Environment Simulator (DES) experiments were conducted so that the pilots performed a tracking task through a central head up display while the direction of the applied acceleration was varied. By altering the position of the centrifuge cab on the arm the direction of acceleration was varied from vertically downwards (+Gz), to lateral to the left (+Gy), lateral to the right (− Gy), and alternately left and right (± Gy, PMGy).

In addition, the degree of physical restraint inside the cab was altered. The shoulder straps were fastened and the inertia reel was unlocked (UNLOCK), the inertia reel was then locked (LOCK) and additional shoulder pads (PADS) were placed just outside of the shoulders to reduce side ways movement under lateral acceleration still further. The first DES experiment examined these three variants under the different acceleration directions.

Heart rate - Figure 5 shows how the heart rate varied with acceleration direction and restraint. As the restraint was improved (UNLOCK - LOCK - PADS) the heart rates fell considerably. Note that the extent of the heart rate change is very much greater than that in Figure 1, where speed was simulated in a fixed base simulator.

Moreover, as the acceleration was imposed to the left (+ Gy), the subjects were forced away from the tracking task display and controls and greater effort was required to bring the head and body back to the center of the cab. This effort produced high heart rates with all degrees of restraint. Lateral acceleration to the right permitted the subjects to use the right arm rests adjacent to the side arm controls, thus slightly less effort was needed and slightly lower heart rates were seen. As the acceleration alternated from right and left, the varying direction required less sustained effort and produced lower heart rates still. When the acceleration was vertical (+ Gz), no effort was required to acquire the display and controls and heart rates were lowest.

The heart rate variance, R wave and variance graphs do not show such clear cut changes as does heart rate. For these three variables, neither acceleration direction nor degree of restraint produced significant changes.

The second DES experiment (II) also examined variants of restraint but the unlock mode was not used. Alternating lateral acceleration, as would be experienced during tracking in the AFTI airplane, was used rather than sustained acceleration.

Heart rate - Figure 6 shows the data from two subjects both with and without PADS. Data are presented for alternating lateral acceleration (PMGy) together with pre and post acceleration controls (static, S and Baseline, B). The figure shows that without the PADS, the heart rates were significantly elevated by lateral acceleration. With the PADS this pattern was not seen; the heart rates gradually declined over the course of the experiment.

Heart rate variance - Figure 7 also shows elevation during lateral acceleration without the pads. In subject M, the PADS graph shows an increase of heart rate variance with acceleration but the pattern is not so obvious for the NO PADS condition.

R wave - Figure 8 shows no deviations with acceleration in either the PADS or NO PADS condition.

R wave amplitude variance - Figure 9 shows significant changes with acceleration in the NO PADS condition while the PADS curves show no changes.

The graphs for EMG (total voltage and center frequency) show no changes with acceleration. The tracking task data (recorded as integrated error for a single pursuit task sequence) do show that tracking was worse during the acceleration, particularly so when shoulder pads were absent (Fig. 10). Tracking during the control periods was similar for both conditions.

DISCUSSION

Heart rate has long been used as an index of arousal, effort, and workload. Many factors can affect heart rate but if these are controlled and only physical or mental effort is altered, their effect on heart rate can be examined. In the fixed base simulator studies, the subjects were well rested, the level of arousal was constant and little physical effort was required. Varying the speed of the simulated aircraft and targets required more accurate tracking and faster decisions. This increased the heart rate of all the subjects. Although the increases were small, they were statistically significant ($p < .05$ and .001) and present in all the tasks examined.

Heart rate increases were also seen in the DES experiments. Here the mental workload was constant, the task did not change, neither did the acceleration level, only the direction and restraint. When sustained lateral acceleration was imposed, the subjects were forced away from the display and controls and considerable physical effort was required to reacquire them. This produced increases in heart rate but of a level considerably higher than those in the fixed base simulator. This difference is a measure of the differential response to physical and mental workload. The worsening of restraint from the PADS to the UNLOCK mode meant that the subjects had to work even harder. Similar results were seen in the second DES experiment. Alternating lateral acceleration required slightly less physical effort when the shoulder pads were fitted and there was little change of heart rate with acceleration. However, without the pads, acceleration increased the heart rate.

The use of heart rate variance as a measure of workload is of relatively recent origin. The literature on the changes of heart rate variance with workload is conflicting. In these studies, heart rate variance rose as speed increased and as lateral acceleration and poor restraint were introduced. However, the changes were not as consistent as those of heart rate itself.

The R wave of the EKG complex is associated with venticular function. High amplitudes signify increased cardiac output and greater cardiac effort. Normally, as heart rates increase, the R wave should decline in amplitude as individual heart beats eject smaller volumes of blood. The more frequent beats, however, contribute to greater cardiac output in unit time. In these studies, R wave amplitude declined with increase in speed and heart rate but showed little change with acceleration. The variance of the R wave, however, was most useful, it rose with speed and acceleration, significantly so when the restraint was decreased.

From these studies it can be seen that the combination of increases in heart rate, heart rate variance, and R wave variance denote an increase in workload (physical, mental or combined).

The EMG data were disappointing. Laboratory studies show that the EMG can indicate muscular effort and fatigue by examination of the total voltage and center frequency. However, laboratory studies involve sustained muscular effort up to 75% of maximum and for long periods inducing severe fatigue. In these studies, the muscular effort was slight and short periods only were required, so it is not so suprising that no significant results were seen in the EMG data. Furthermore, the electrodes and leads were prone to interference and while the electrodes were non-invasive (stick-on skin electrodes) the leads on the arm and neck did prove cumbersome. Future flight studies will not involve EMG data collection.

Tracking effort has been used repeatedly in the past to measure pilot performance. Good tracking increases the times for and chances of hits on targets. In these studies, the electrical signals recorded as tracking failed to show significant results. However, the integrated tracking scores logged by hand did show that lateral acceleration decreased the scores, particularly so when restraint was reduced. It may be that the signals recorded from the simulator computer were less accurate than the integration of the computer itself.

A major limitation of this study has been the number of subjects. Clearly, it is wise to restrict access to such a valuable airplane as the AFTI F-16 but this has meant using a small, highly-trained group of subjects who clearly cannot represent the pilot population as a whole. However, in the event it has proved possible to derive some measures of physiology and performance that can indicate workload changes and these will be used in future flight studies to compare flight regimes, systems, and procedures for workload changes. Furthermore, the individual data recorded from each subject has contributed to the data bank that can be used when those subjects perform in flight tests. The greater value of the data recorded may well appear in later studies.

CONCLUSIONS

Elevations of heart rate, variance, and R wave variance, especially when simultaneous, indicate conditions of increased workload for the subjects used in this study. This should be true for inflight studies also.

In lateral acceleration, shoulder pads are of considerable value. They reduce the workload, the effort, and improve tracking. The improvement in heart rate alone was some 30 beats/min. This should significantly delay the onset of pilot fatigue and their physical limitations. Shoulder pads should be fitted to the AFTI F-16 and all other aircraft which exhibit sustained lateral acceleration.

EMG data are not so reliable in simulated flight. This combined with the additional electrodes and leads and their position on the body all militate against comfort and efficiency and will be eliminated in the flight studies.

Tracking data will be recorded in the flight phases of the AFTI F-16 program. In the hope that more accurate data may show significant effects in the future, analyses will be continued to check whether these data can be used to measure pilot workload.

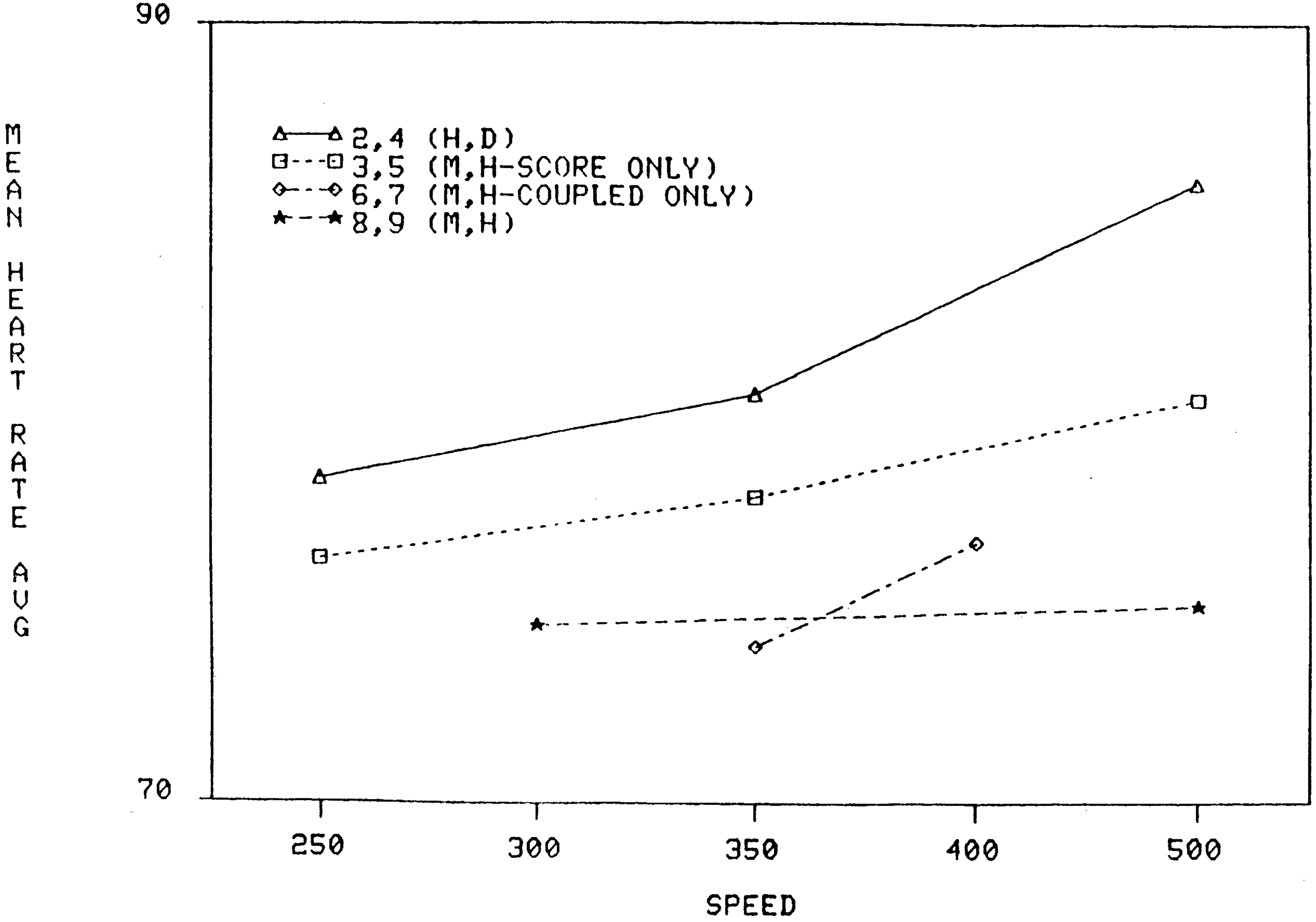


Figure 1 Mean Heart Rate Average and Speed

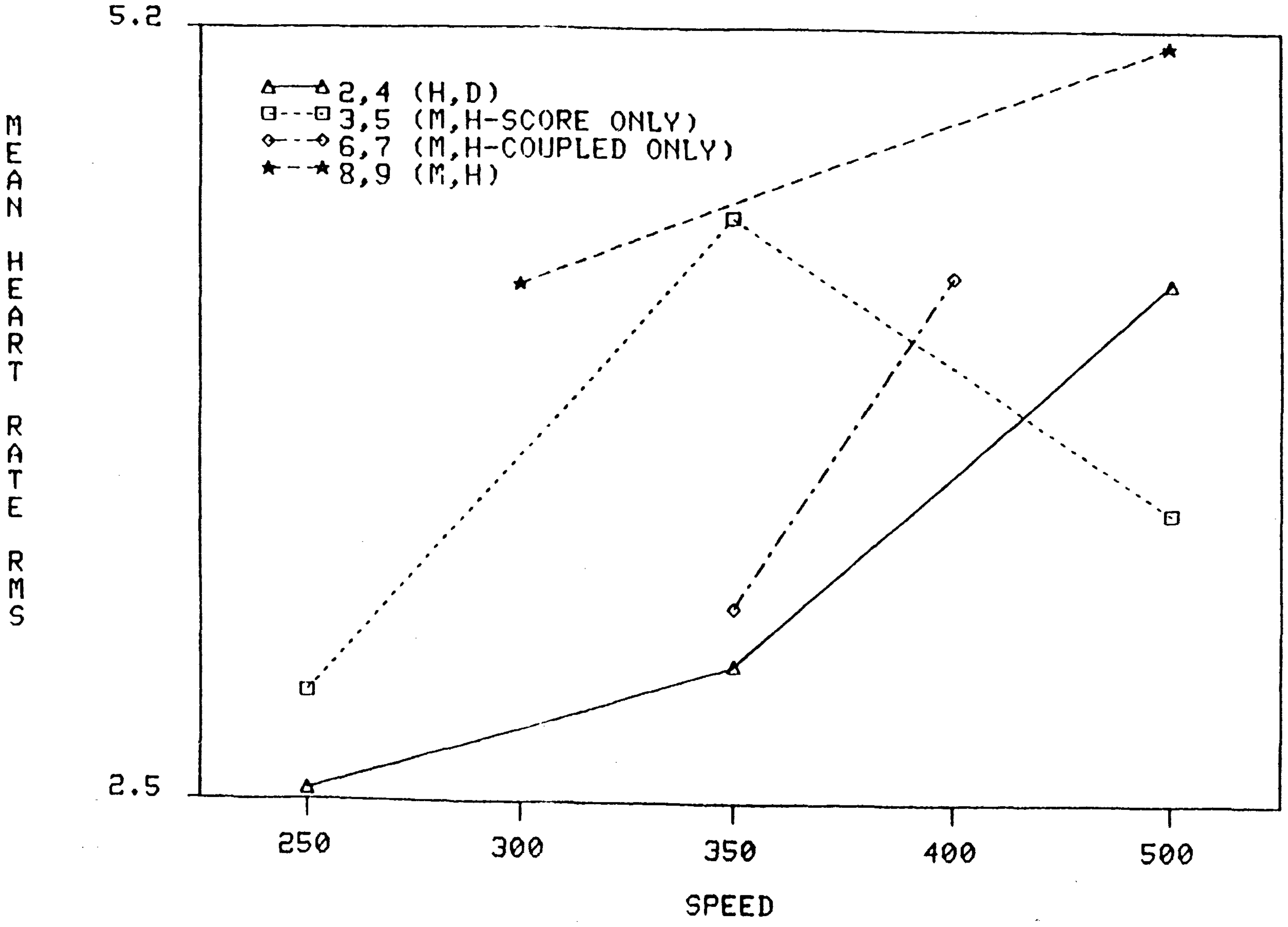


Figure 2 Mean Heart Rate Variance and Speed

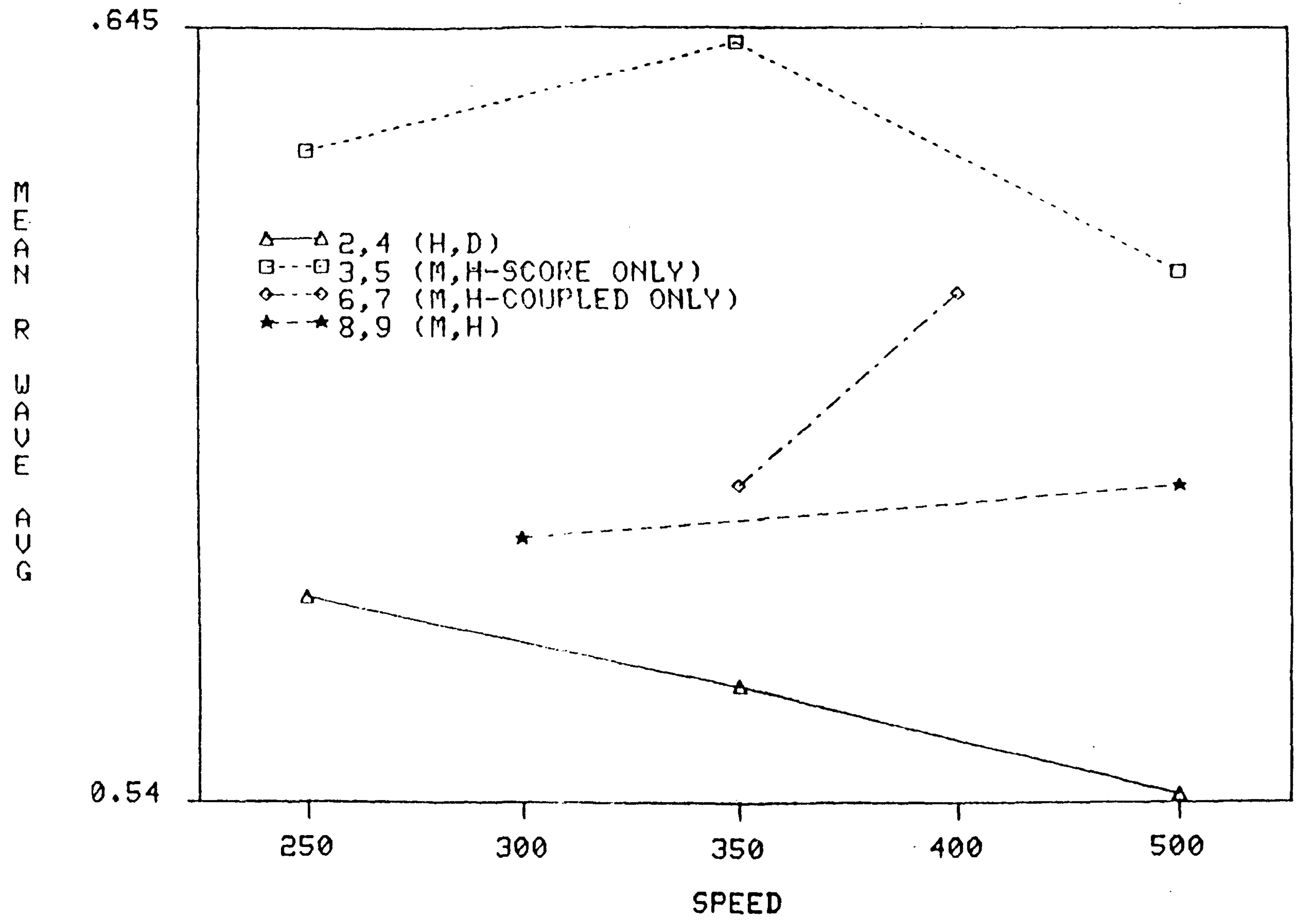


Figure 3 Mean R Wave Average and Speed

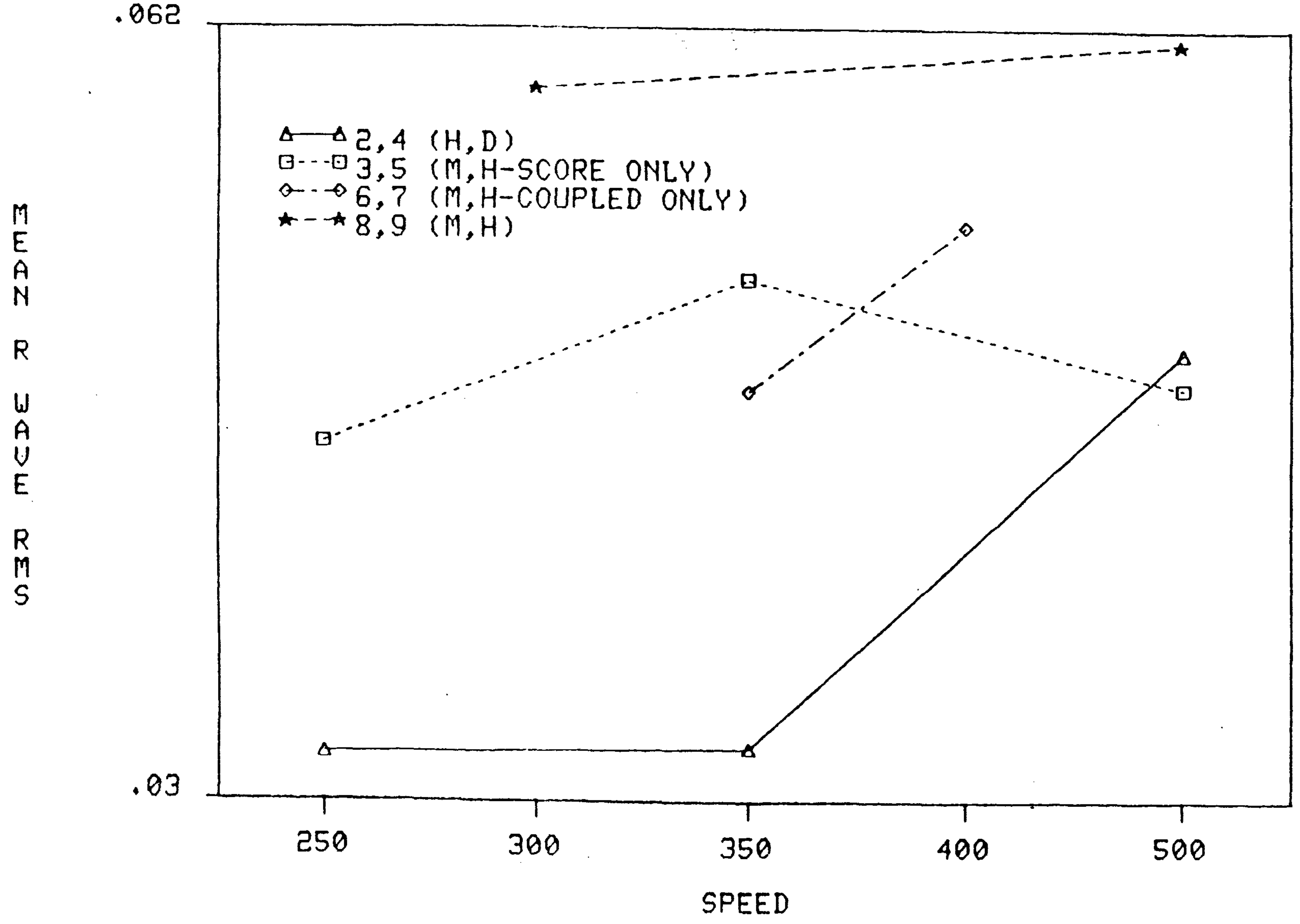


Figure 4 Mean R wave Variance and Speed

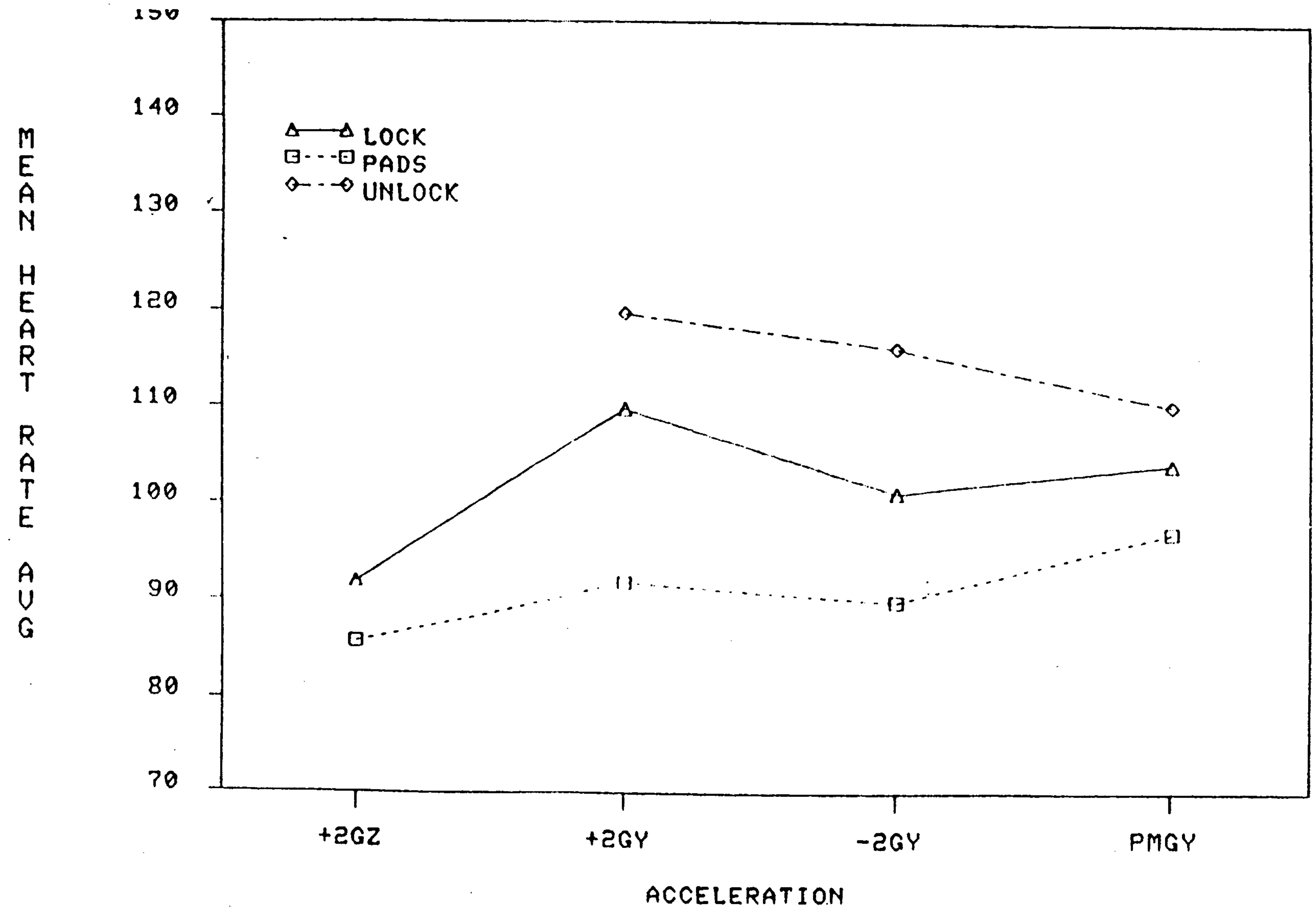


Figure 5 Mean Heart Rate Average and Acceleration I

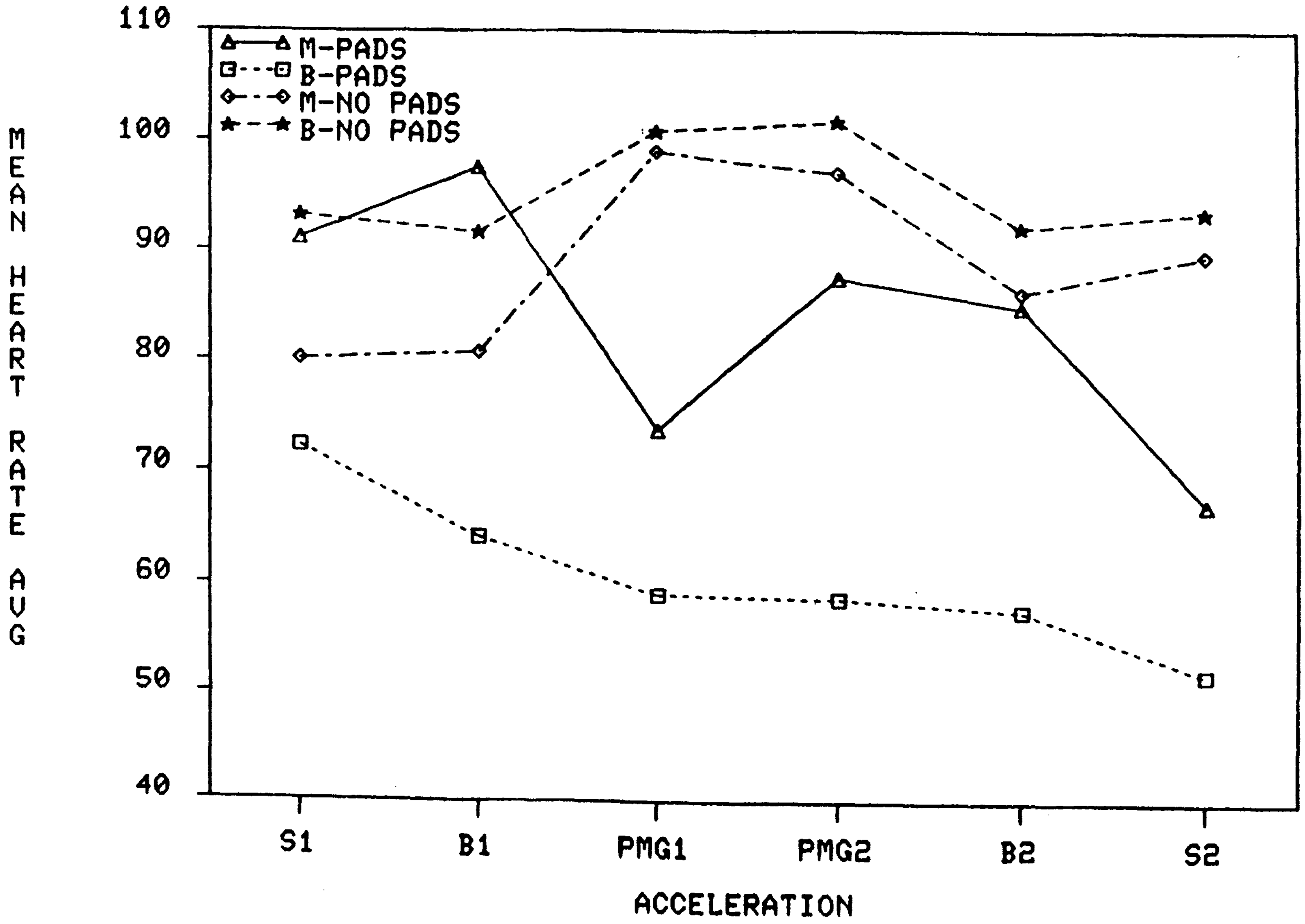


Figure 6 Mean Heart Rate Average and Acceleration II

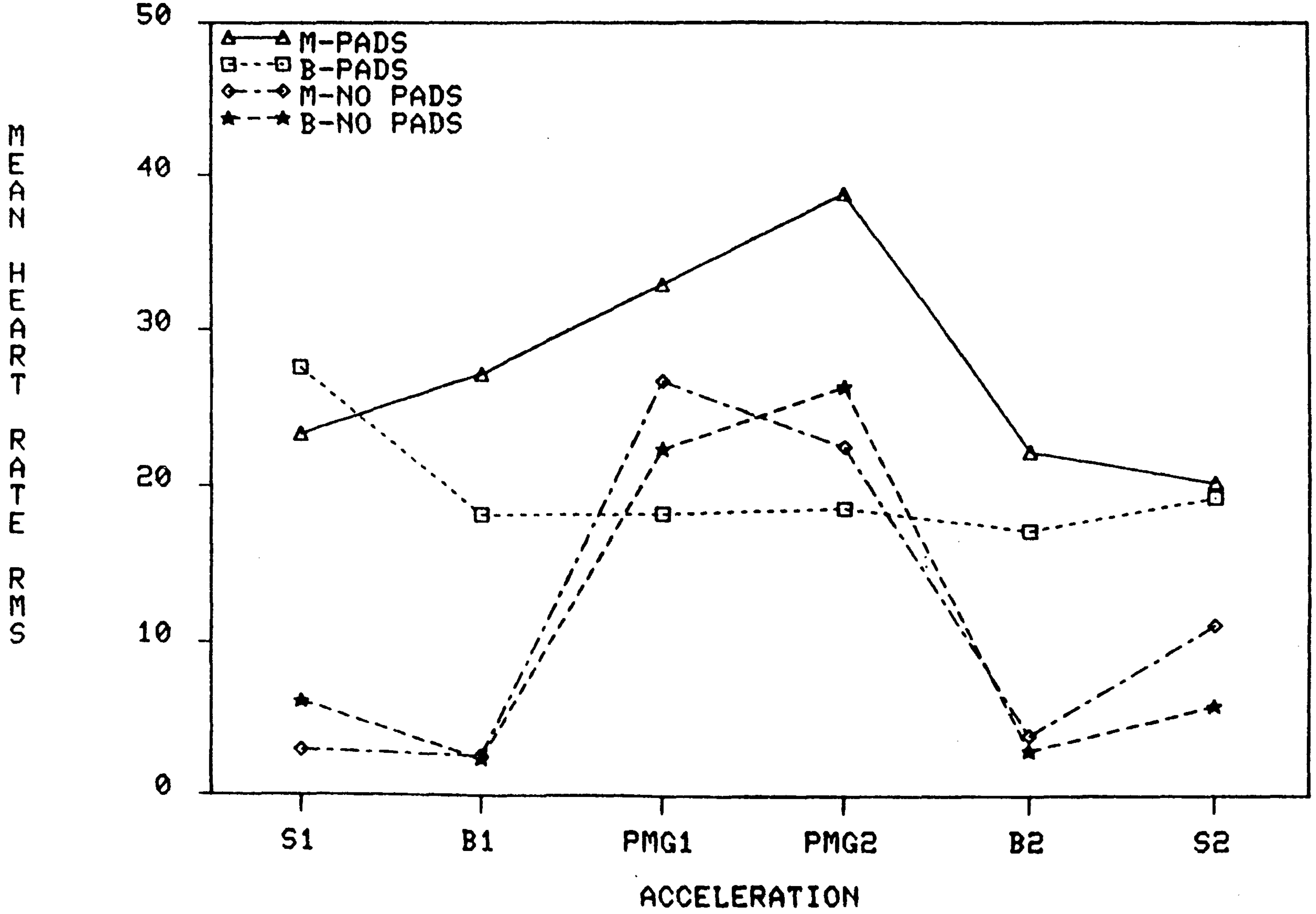


Figure 7 Mean Heart Rate Variance and Acceleration II

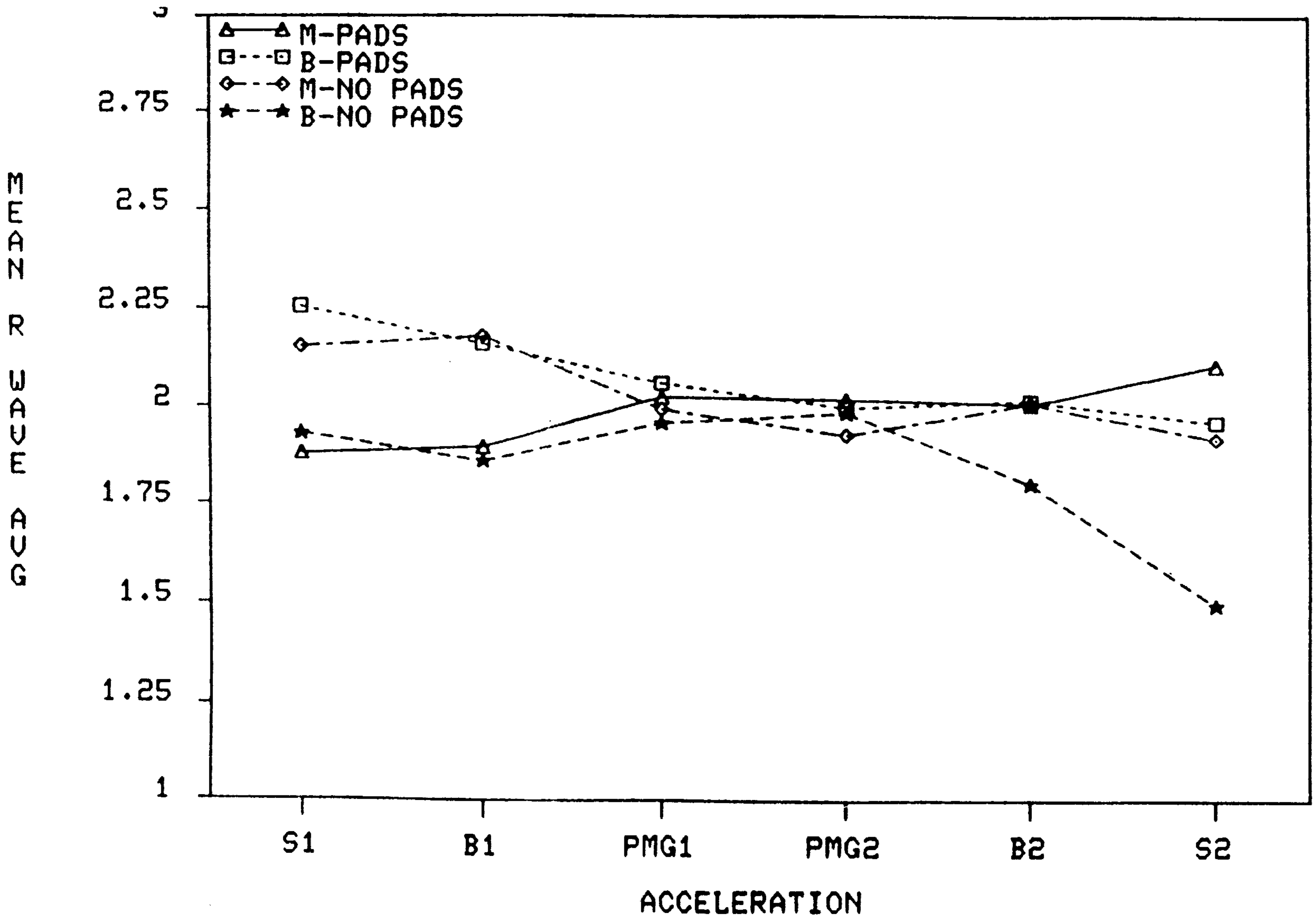


Figure 8 Mean R Wave Average and Acceleration II

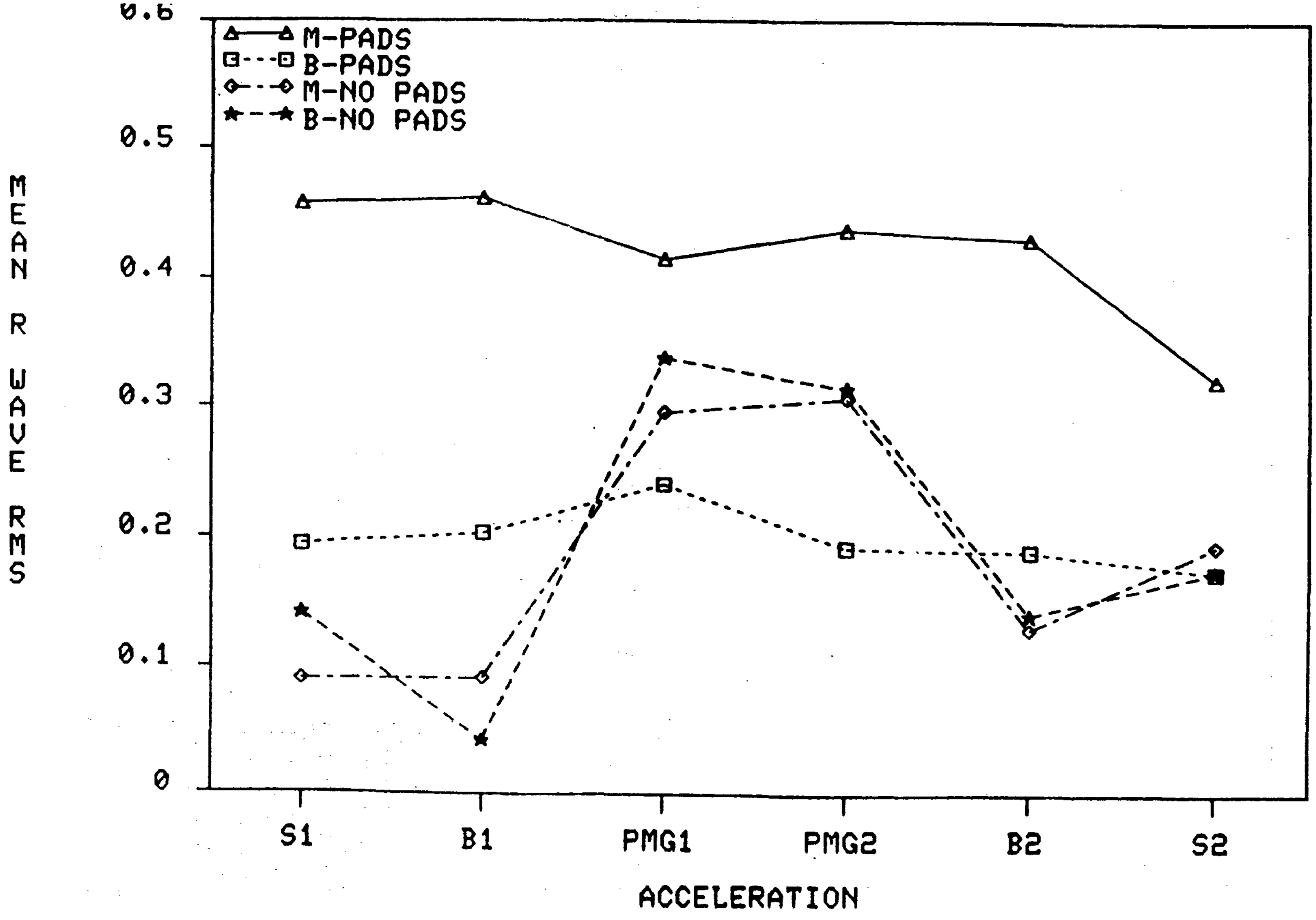


Figure 9 Mean R Wave Variance and Acceleration II

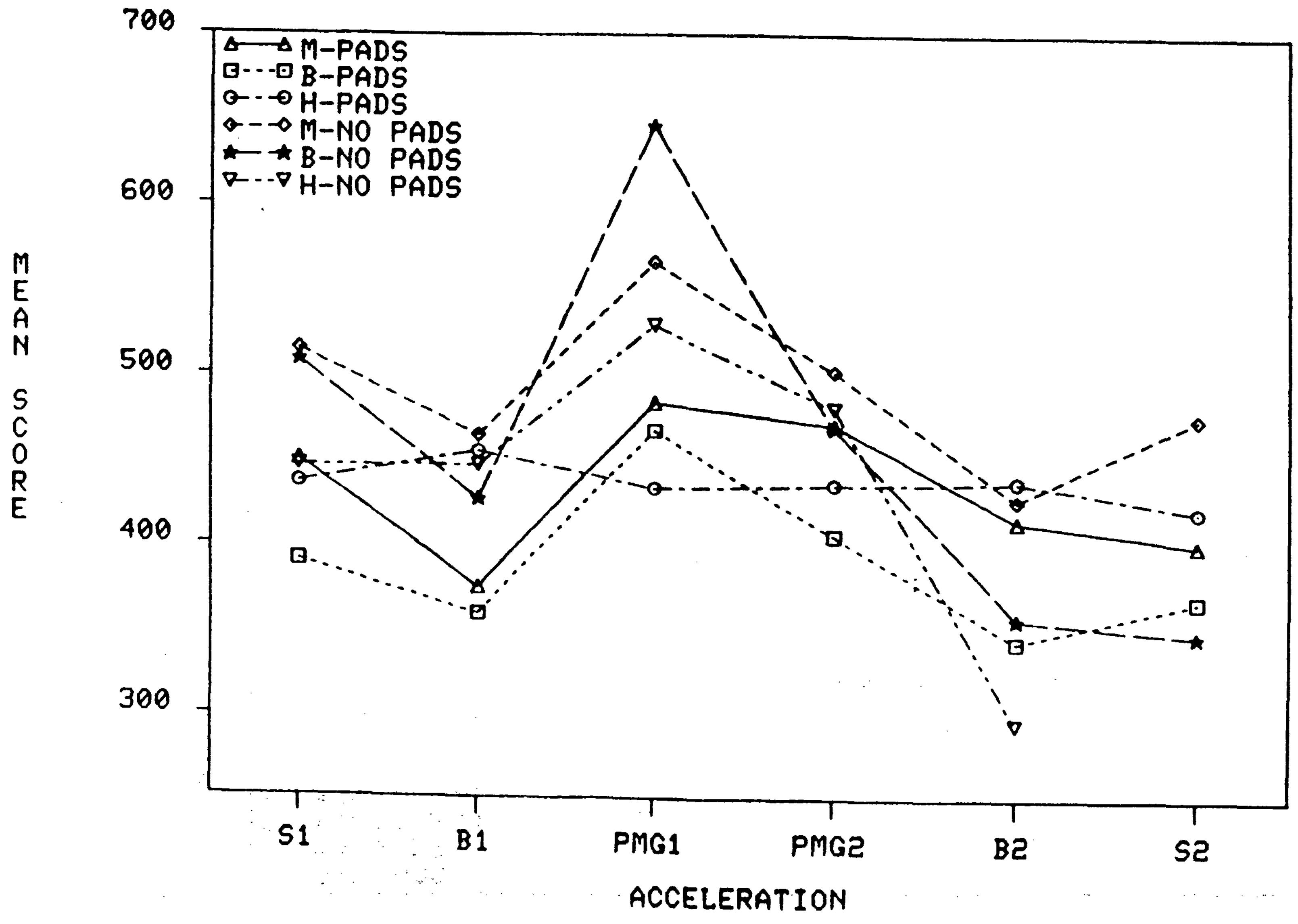


Figure 10 Mean Score and Acceleration II

HEART RATE AS AN IN-FLIGHT MEASURE OF PILOT WORKLOAD

Alan H Roscoe

Royal Aircraft Establishment, Bedford, England

HEART RATE AS AN IN-FLIGHT MEASURE OF PILOT WORKLOAD

Alan H Roscoe

Royal Aircraft Establishment, Bedford, England

Monitoring heart rate in flight is a relatively simple procedure; the technique is not intrusive, it does not compromise flight safety, the signals are easy to record, and the discrete nature of the data make them amenable to various forms of analysis. It is not surprising, therefore, that a large number of experiments have been reported in which this physiological variable has been recorded in flight. Although most of these experiments were designed primarily to examine the effects of various physical and mental stressors on pilots a small number was aimed specifically at estimating levels of workload (1), (2) and (3). There is now unequivocal evidence that pilots heart rates tend to increase during flight and especially during such demanding manoeuvres as the take-off and the landing.

Using heart rate to estimate workload in this way prompts one to ask a number of questions:

1 What is the relationship between a pilot's heart rate and his workload?

2 Is heart rate a valid and reliable indicator of workload?

3 If it is - how should it be used?

4 What are the likely neuro-physiological mechanisms involved?

These questions will be discussed using examples of heart rate selected from more than 3000 plots recorded during flight trials at RAE Bedford. But first it is important to describe what is meant by the term pilot workload. There are many definitions of workload most of which appear to fall into 2 broad conceptual areas, those that relate to the task or to the demands of the task and those that are associated with the response or effort. In this paper workload is considered to be related to effort, an interpretation which is compatible with the use of physiological variables and which also lends itself readily to subjective assessment. In this context it is worth noting that some 80% of pilots view workload as being effort-related (4), a view which agrees well with the influence on the piloting task of such individual factors as natural ability, response to stress, physical fitness, age, training and experience.

1 THE RELATIONSHIP BETWEEN HEART RATE AND WORKLOAD

The most used and probably the most reliable methods of estimating workload in flight are those based on some form of subjective reporting by experienced test pilots. And so it is of interest to examine the relationship between pilots assessments of workload and their heart rate responses.

Following a 3 year exploratory study, in which heart rates were recorded from pilots flying a wide variety of aircraft, it was decided in 1972 to monitor heart rate routinely during a series of flight trials to evaluate different types of reduced noise landing approaches (5) and (6).

The first flight trial used a twin turbo-prop HS Andover to compare a number of different approach profiles using a conventional 3° glide slope as a datum. Single-segment approaches with gradients of 6°, 7½° and 9° and two-segment approaches with a 7½° slope changing to 3° at 200 ft were studied in detail. Figure 1 shows the senior project pilot's 30 sec heart rates for the single-segment profiles recorded during one of a group of 4 sorties. The experimenter used a Latin square design to allow a realistic comparison to be made. Overall mean heart rates for the 4 approach profiles are shown in Figure 2 for the same pilot. In this case there was exceptionally good agreement between heart rate levels and subjective estimates of workload; and also with expected levels of task difficulty - the workload being expected to increase with steeper approach paths and higher rates of descent.

Later, two-segment approach profiles, with a 7½° slope changing to a 3° slope at 200 ft, were evaluated. Figure 3 compares mean heart rate values for these 2-segment approaches and landings with those for 3° approaches and landings. Interestingly, despite their relatively low heart rate responses the project test pilots initially rated the workload for the two-segment approaches as high. It later transpired that these 2 pilots had instinctively disliked the idea of changing from a steep gradient - with the higher rate of descent - to a normal gradient at 200 ft. After the first sortie they modified their views and then consistently rated the 7½°/3° approach as being as easy as, if not easier than, the normal 3° approaches. This example highlights the possibility of subjective assessments of workload being biased by allowing instincts and misconceptions to influence judgement. It also illustrates the advantage of using heart rate to augment - or sometimes to question - subjective assessments of workload.

In a later trial in this series a VC-10 four-jet transport was used to evaluate 5°/3° two-segment approaches - the transition from the steep to the normal gradient being increased to 500 ft for this larger aircraft. Figure 4 illustrates beat-to-beat heart rates recorded from the handling pilot and from the co-pilot during an early two-segment approach and landing. The introduction of beat-to-beat or instantaneous heart rate plots increased the value of this physiological measure by recording short term changes in rate which can be used to identify changes in levels of workload. For example, in Figure 4 'A' indicates the start of descent on the 5° glide slope - in this case at a greater height than usual. Points 'B' and 'C' indicate, respectively, the outer marker and the transition at 500 ft. This type of presentation also provides a bonus measure in the form of sinus arrhythmia.

Figures 5 and 6 compare overall mean 30 sec heart rates for 5°/3° and 3° approaches and landings. These responses confirmed the pilots subjective assessments that the two-segment profiles generated similar levels of workload to the conventional 3° profile.

These examples are typical of flight trials in which different experimental workload levels can be compared in a realistic way with a convenient datum or with each other. Throughout the series there was a substantial measure of agreement between relative workload levels as judged by pilots subjective estimates and by their heart rate values. Such comments made in later discussion as "...the way in which my heart rate consistently increased at that point reflects exactly how I felt about the difficulty ..." and "I was aware of beginning to work harder at that stage of the approach indicated by an appreciable increase in my heart beat ..." are typical.

A number of other flight trials at Bedford has resulted in similar levels of agreement between subjective estimates of workload and heart rates. For example, in trials to assess the value of a "ski-jump" take-off technique for Harrier VSTOL aircraft heart rate responses agreed with pilot ratings that workload levels for these take-offs were probably less and certainly no greater, than those for conventional short take-offs in this aircraft (7).

During a recent series of flights to evaluate economic category 3 landings - consisting of autopilot approaches to a 50 or 60 ft decision height and then a manual flare and touchdown (8) - pilots heart rates and workload ratings (using a 10-point scale) for the decision and landing were recorded. Figure 7 is a typical beat-to-beat heart rate plot showing the rapid increase in workload as decision height was neared and manual control was assumed for the landing. The scatter diagram (Figure 8) illustrates graphically the relationship between 32 heart rate responses and workload ratings in real fog for the senior project test pilot. These data varied more or less according to fog conditions and runway visual ranges (RVRs).

Unlike the noise abatement trials workload levels during fog approaches and landings could not be compared directly with a suitable standard. Nevertheless, heart rates recorded in fog could be compared indirectly with those recorded during approaches and landings in clear weather. Pilots subjective estimates of workload in fog tended to be based partly on comparison with those in clear weather and - using the rating scale - on an awareness of the degree of spare capacity available for other tasks (Figure 9). There was also a tendency for pilots to compare workload levels on different approaches during the same sortie as fog conditions varied.

Flight trials such as these have appeared to provide strong evidence of a reasonably good relationship between a pilot's heart rate response and his estimate of the workload level associated with a well defined and demanding piloting task. And it is a relationship that appears to hold good both for comparative levels of workload and for short term changes in workload as indicated by changes in beat-to-beat heart rate.

Unfortunately, when dealing with human subjects - even with experienced test pilots - discrepancies and inconsistencies are bound to occur between their opinions and their heart rate responses. In most such instances at Bedford a plausible cause for the disagreement has been identified.

2 HEART RATE AS AN INDICATOR OF WORKLOAD

The use of physiological variables to indicate levels of workload has been viewed with suspicion by many people and the use of only one variable - such as heart rate - has been criticised in particular. However, many of these criticisms have been based on the results of laboratory and flight simulator experiments where often the task and levels of workload were unrealistic.

Experience at Bedford has shown that when the pilot is in the handling loop - or expecting to enter the loop, and when the flight task is reasonably demanding heart rate alone will usually identify meaningful changes and differences in workload. Of course, expected changes in workload may be more theoretical than practical; and so before deciding whether heart rate can differentiate between workload levels it is important to be sure that there is, in fact, a real difference (7).

When the flight task is relatively undemanding or when the pilot is in a purely monitoring role heart rate *per se* may not differentiate between small differences in workload. But often in these instances visual inspection of beat-to-beat plots will reveal changes in the degree of heart rate variability (sinus arrhythmia) which may well signify changes in workload (9).

3 USING HEART RATE TO ASSESS LEVELS OF PILOT WORKLOAD IN FLIGHT

When monitoring pilots in flight it is obviously desirable to obtain their active cooperation and it is even better to have their enthusiastic support for the technique. At Bedford, test pilots frequently apply their own electrodes before flight; and at some time afterwards it is quite usual for them to express a keen interest in the recorded data so that they can relate their subjective impressions of workload and task difficulty to changes in their heart rate.

Heart rate indicates only relative differences in workload and so it is helpful to have some form of datum for purposes of comparison. In practice assessment of workload is usually associated with the introduction of a new aircraft system or operating technique and so one can often compare the new with the old. Although it is not always possible to compare heart rate responses for different experimental variables during the same sortie, or even under similar flight conditions, the advantages of doing so are obvious.

The individual nature of heart rate responses make it almost essential, especially when dealing with small numbers of pilots, for each pilot to be considered as his own control.

A pilot may compensate for an easier task by improving his performance or, conversely, he may allow his performance to deteriorate rather than exert more effort to meet the demands of a more difficult task. In each case his workload - and thus his heart rate - may remain unchanged; and so it must be axiomatic that when assessing workload performance criteria are clearly defined and monitored.

As mentioned earlier differences in workload are more likely to be detected by heart rate and probably by subjective assessment when the task is realistically demanding. And so the technique is particularly appropriate for estimating workload during the approach and landing. In this instance the task is well defined and performance can usually be monitored by on-board instrumentation and by airfield-sited kinetheodolites or radar.

The high cost of operating research aircraft usually makes it impossible to obtain enough data for worthwhile statistical analysis. Nevertheless, obvious trends in heart rate changes can be used in conjunction with pilot ratings to provide valuable and reliable indications of differences in workload levels. Surprisingly, despite being more used to obtaining precise measurements from mechanical and electronic devices, trials scientists at Bedford have found pilots heart rate levels and subjective ratings to be of definite value for assessing or comparing levels of workload in flight.

4 POSSIBLE NEURO-PHYSIOLOGICAL MECHANISMS

There is a substantial amount of evidence in favour of workload being the main determinent of heart rate levels in experienced pilots during demanding flight manoeuvres. It is interesting to speculate on possible neuro-physiological mechanisms that would explain this relationship. Certainly, it is rarely due to physical activity - which during normal flight is very low. Although the fact that heart rates are higher for pilots in the handling loop does suggest that increased neuromuscular activity of some form must play a part.

Piloting an aeroplane, especially during the more difficult manoeuvres, requires the brain to collect, filter and process information quickly; to exercise judgement and make decisions; and to initiate rapid and appropriate actions. This neurological activity - which must have been essential for the survival of primitive man - is associated with a state of preparedness sometimes known as arousal. Furthermore there is experimental evidence that increased arousal - up to a moderate level - enhances a person's capacity for complex skills and thus improves performance (10). For instance, Duffy (11) reported that the degree of arousal "... appears to affect the speed, intensity and coordination of responses and thus to affect the quality of performance". She also observed that in general the optimum level of arousal appears to be a moderate level with the curve expressing the relationship between performance and arousal taking the form of an inverted 'U'. Other authors have also referred to such a relationship though there is only meagre experimental evidence to support it (12). Nevertheless, a theoretical relationship of this type has a particular attraction in the context of flying aeroplanes as there is evidence that both under- and over-arousal have preceded landing accidents where pilot performance was clearly below an acceptable level.

There is some experimental evidence that a similar inverted 'U' shaped function describes the relationship between performance and task demands (13). And it has been suggested that levels of arousal are determined by task characteristics or demands, by how the individual perceives the situation, and by how he responds to his environment (14) (15). And so one can speculate that a pilot, by matching his level of arousal to the perceived difficulty of a flight task, is more likely to produce an adequate - if not optimum - level of performance. The result will depend largely on his training and experience, although if the task is a novel one, as frequently happens in test flying, a significant element of empiricism must be involved. Clearly, the level of arousal should be high enough for the task *per se* and also high enough to allow for the unexpected. For example, an engine failure on take-off may require extremely rapid and appropriate actions.

On occasions, at Bedford, it has been obvious from the sudden increase in heart rate after the start of a manoeuvre that a pilot had failed to anticipate the difficulty of the task and "tune" his arousal accordingly. Conversely, high heart rates have been recorded - both before and during a manoeuvre - when there was an element of uncertainty about the task. This was particularly noticeable for the novel "ski-jump" take-offs and for a pilot's first approach and manual landing in fog. The probability of a near optimum level of arousal being generated must be greater if a pilot has recently experienced the demands of a particular flight task. Heart rates recorded from several pilots during sorties of approaches and landings in fog became lower and reasonably consistent after the first or second run. Zwaga (16) in describing an "adjustment period" during which physiological responses to specific stimuli were moderated related the phenomenon to the concept of arousal.

Anticipatory increases in heart rate seen before the start of manoeuvres such as the take-off presumably indicate an increased level of preparedness in the pilot's neuro-physiological system for the demanding task to follow. It seems clear that increasing the level of arousal in this way must be an advantage in the same way that sportsmen "warm up" before competitive events. In other words, a task requiring a high level of psychomotor skill should not be started from cold.

Support for these speculations is provided by experimental evidence showing that appropriate pathways in the brain and central nervous system do exist. Stimulation of the reticular activating system (RAS) results in increased alertness, improved information processing, and shorter reaction times. This state of increased arousal is apparently sustained by reciprocal feedback mechanisms between the cortex, the RAS and the hypothalamus. The hypothalamus, in addition to regulating autonomic nervous system activity - which includes heart rate, also contains integrating and organising centres concerned with arousal.

Although the concept of arousal is an oversimplification of complex neuro-physiological mechanisms it is functional and, providing it is not confused with emotion, it conveniently explains the relationship between a pilot's workload and his heart rate.

A final example from Bedford may confirm - in a practical way - the importance of an adequate level of arousal during the approach and landing. Figure 10 shows two beat-to-beat heart rate plots for the same pilot during a sortie of 6^{o} approaches using direct lift control (DLC) in a BAC 1-11. The upper trace was recorded during the ninth approach which ended in a particularly heavy landing when the pilot failed to arrest the rate of descent. Damage to the aeroplane necessitated grounding for three weeks. Uncharacteristically the heart rate did not increase as the runway threshold was neared - although it increased rapidly after the hard touchdown! (The temporary interference in the trace was caused by the jolt on landing affecting the on-board monitoring equipment). The lower trace, of a typical response for a 6^{o} approach and roller landing, is shown for comparison.

CONCLUSIONS

The questions posed at the beginning of this paper cannot be answered with any degree of scientific certainty. Nevertheless, it is possible from the discussions above to arrive at the following conclusions:

1 There is good evidence that heart rate responses increase with increased workload.

2 Differences in heart rate values appear to indicate relative differences in workload.

3 Heart rate monitoring is best used in conjunction with a rating scale for workload.

4 Although the exact nature of the neuro-physiological mechanisms involved is not known it is possible to construct a reasonable hypothesis using the concept of arousal.

Reference was made earlier to the difficulty of defining workload - perhaps in the interest of clear thinking one should avoid using this term altogether and refer instead to pilot activity, effort, task demands and so on.

REFERENCES

1 Roman J et al. Flight research program VI: Heart rate and landing error in restricted fields of vision landings. Aerospace Med 38, 128-132, 1967.

2 Ruffell Smith H P. Heart rate of pilots flying aircraft on scheduled airline routes. Aerospace Med 38, 1117-1119, 1967.

3 Hasbrook A H, Rasmussen P G. Pilot heart rate during in-flight simulated instrument approaches in a general aviation aircraft. Aerospace Med 41, 1148-1152, 1970.

4 Ellis G, Roscoe A H. The airline pilot's view of flight deck workload: A preliminary study using a questionnaire. Royal Aircraft Establishment Technical Memorandum - to be published.

5 Roscoe A H. Heart rate monitoring of pilots during steep gradient approaches. Aviat Space Environ Med 46, 1410-1415, 1975.

6 Roscoe A H. Pilot workload during steep gradient approaches. In: Conference Proceedings No 212 Aircraft Operational Experience and its impact on safety and survivability. AGARD, Paris, 1976.

7 Roscoe A H. Handling qualities, workload and heart rate. In: Eds Hartman B O and McKenzie R E, Survey of methods to assess workload. AGARDograph No 246, 1979.

8 Roscoe A H. Pilot workload and economic category 3 landings. In: Conference pre-prints. Aerospace Medical Association Annual Scientific Meeting, 1980.

9 Roscoe A H. Heart rate changes in test pilots. In: Eds Kitney R I and Rompelman O, The study of heart-rate variability. Clarendon Press Oxford, 1980.

10 Malmo R B. Activation: A neuro physiological dimension. Psychological Rev 66, 367-386, 1959.

11 Duffy E. Activation and behaviour. J Wiley & Sons, New York, 1962.

12 Welford A T. Arousal, channel capacity and decision. Nature 194, 365-366, 1962.

REFERENCES (Contd)

13 Welford A T
Stress and performance.
Ergonomics 16, 567-580, 1973.

14 Wilkinson R T, El-Beheri S, Gieseking C C
Performance and arousal as a function of incentive, information load and task novelty.
Psychophysiology 9, 589-599, 1972.

15 Kahneman D
Attention and effort.
Prentice-Hall Inc, Englewood Cliffe, 1973.

16 Zwaga H J G
Psychophysiological reaction to mental tasks - effort or stress.
Ergonomics 16, 61-67, 1973.

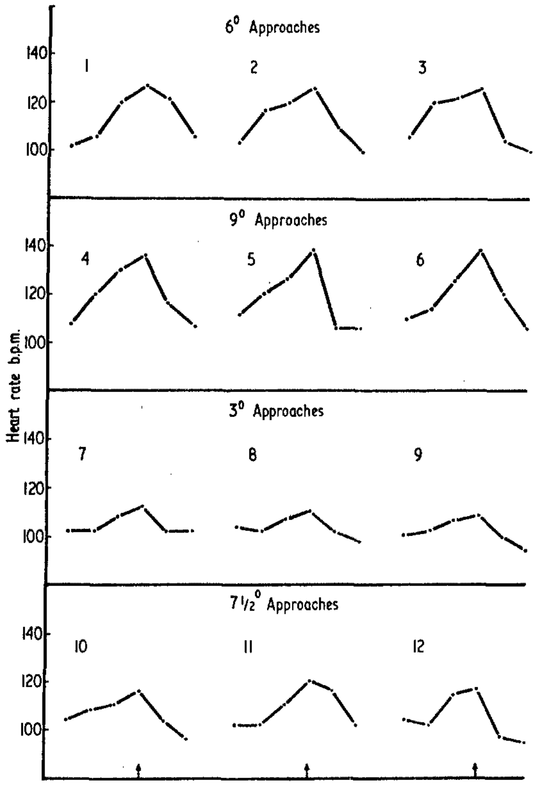


Fig 1 Mean 30 sec heart rates for one sortie of experimental approaches (HS Andover)

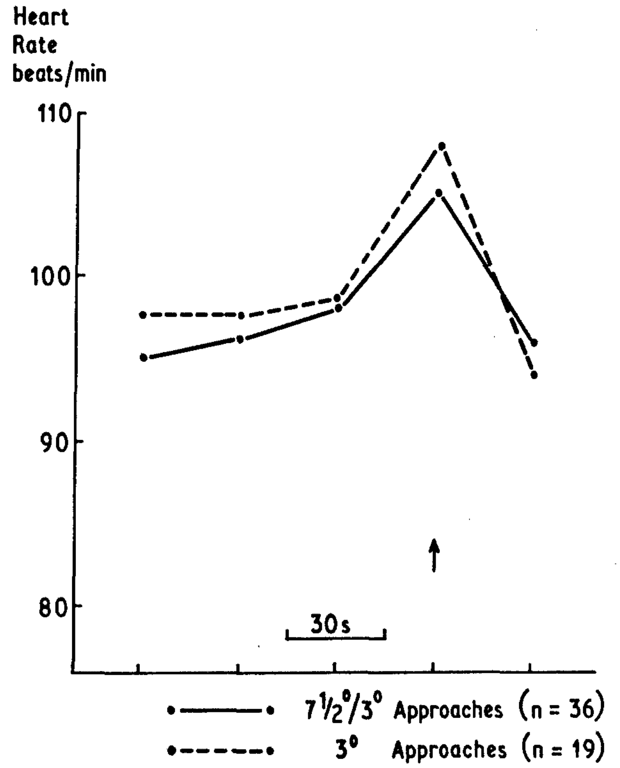


Fig 3 Overall mean heart rates for 7½°/3° and 3° approaches (HS Andover)

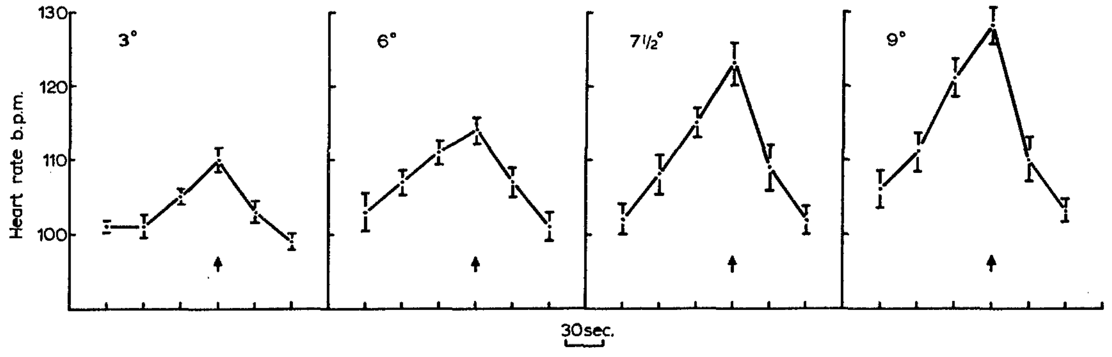


Fig 2 Overall mean 30 sec heart rates (± SE) for four sorties of experimental approaches (HS Andover)

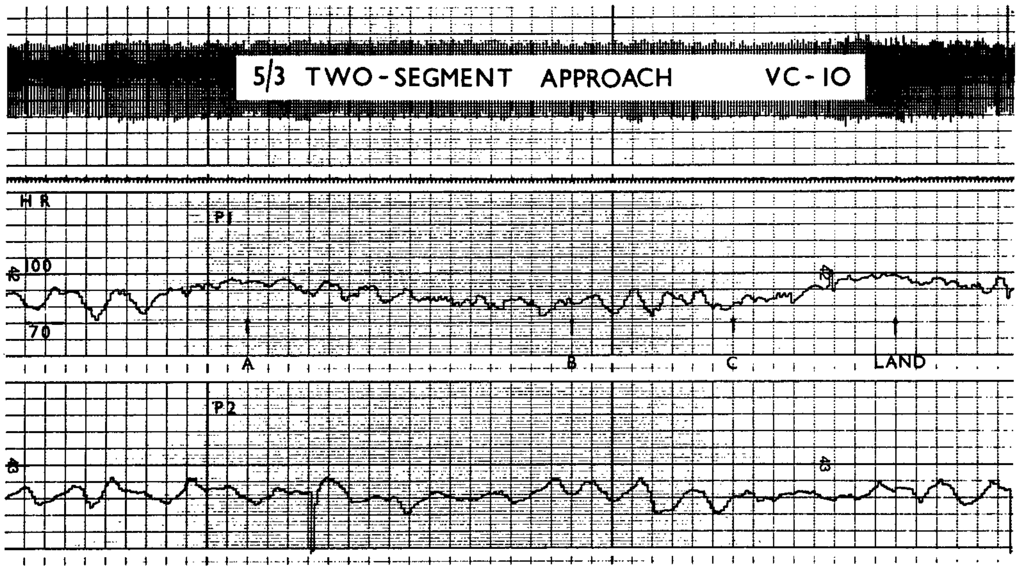


Fig 4 Beat-to-beat heart rate for two pilots during $5^{\circ}/3^{\circ}$ approach

A - Start of descent B - Outer marker C - Transition

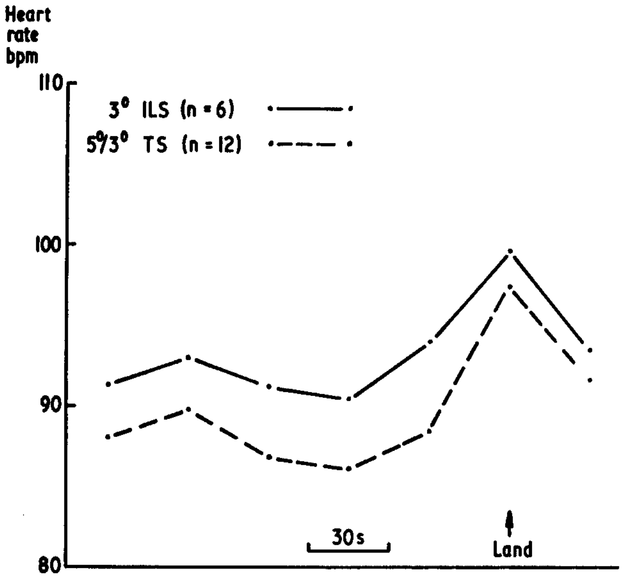


Fig 5 Pilot No 1

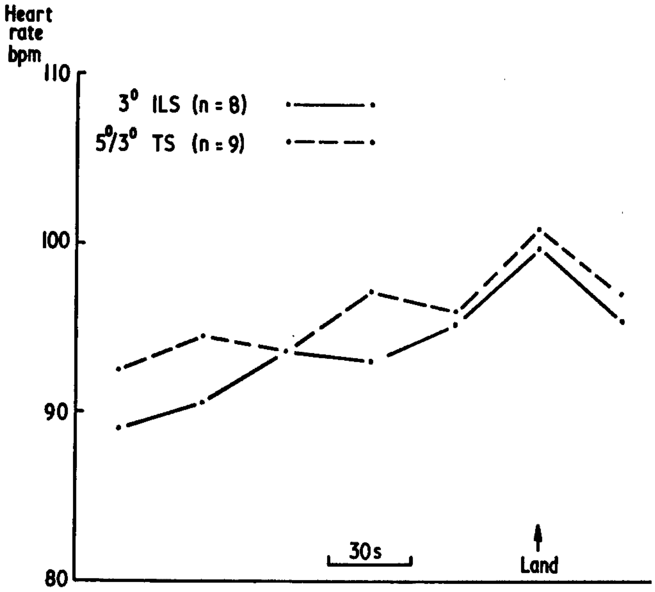


Fig 6 Pilot No 2

Overall mean 30 sec heart rates for $5^{\circ}/3^{\circ}$ and 3° approaches (UC-10)

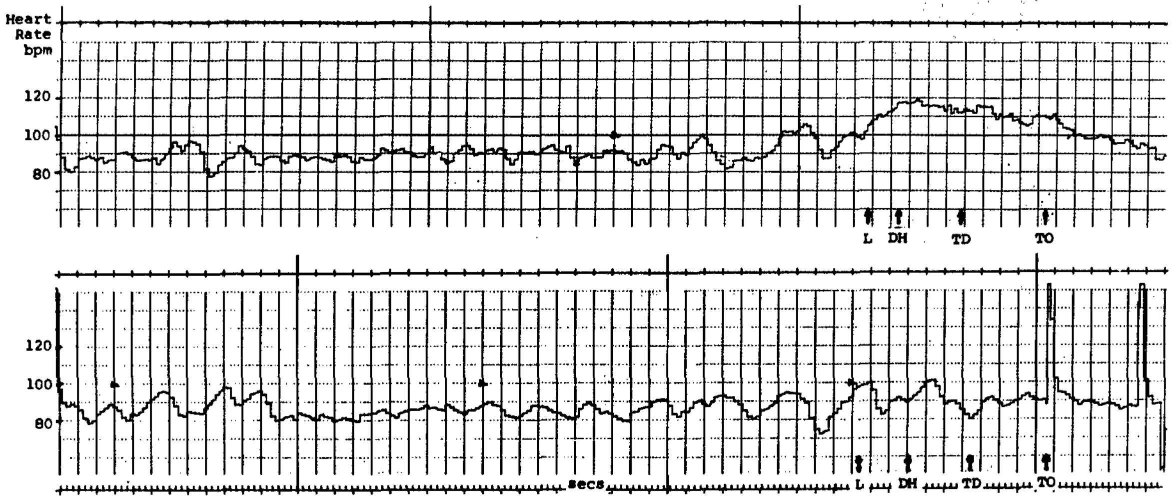


Fig 7 Heart rates during approaches and landings in fog (BAC 1-11)
Upper trace Economic Category 3 with manual landing
Lower trace Category 3 autopilot approach & landing

L - Lights DH - Decision height TD - Touchdown TO - Take-off

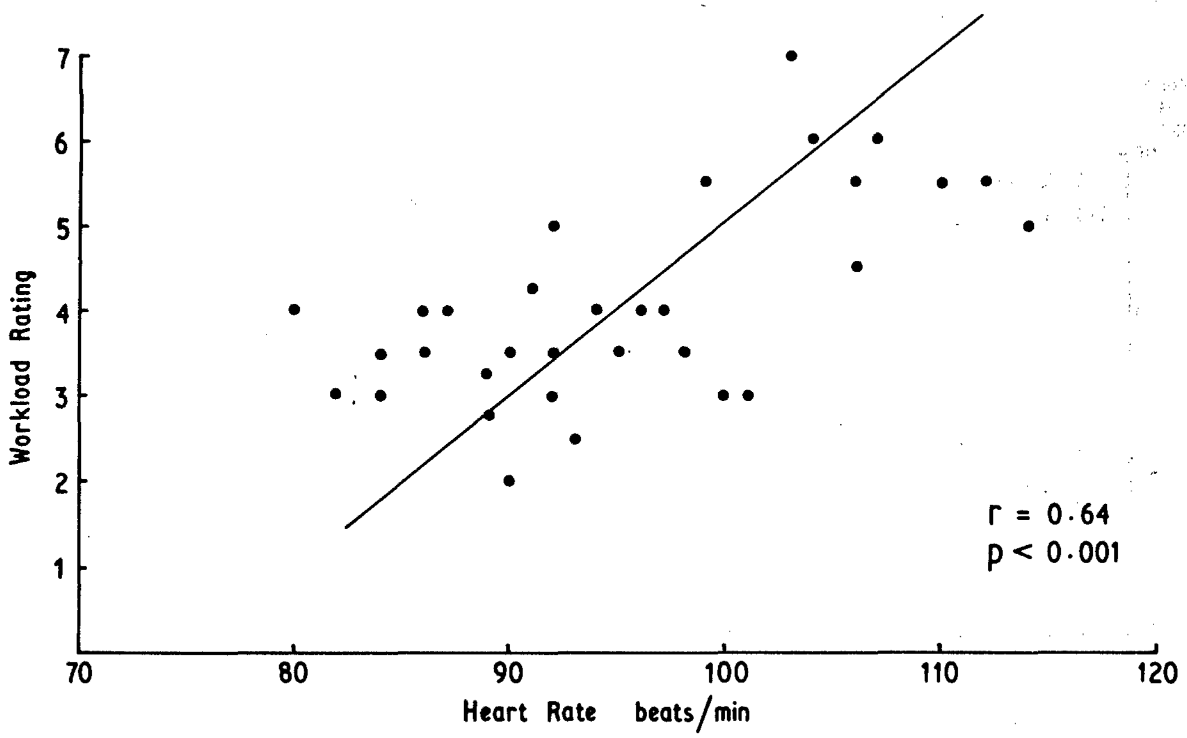
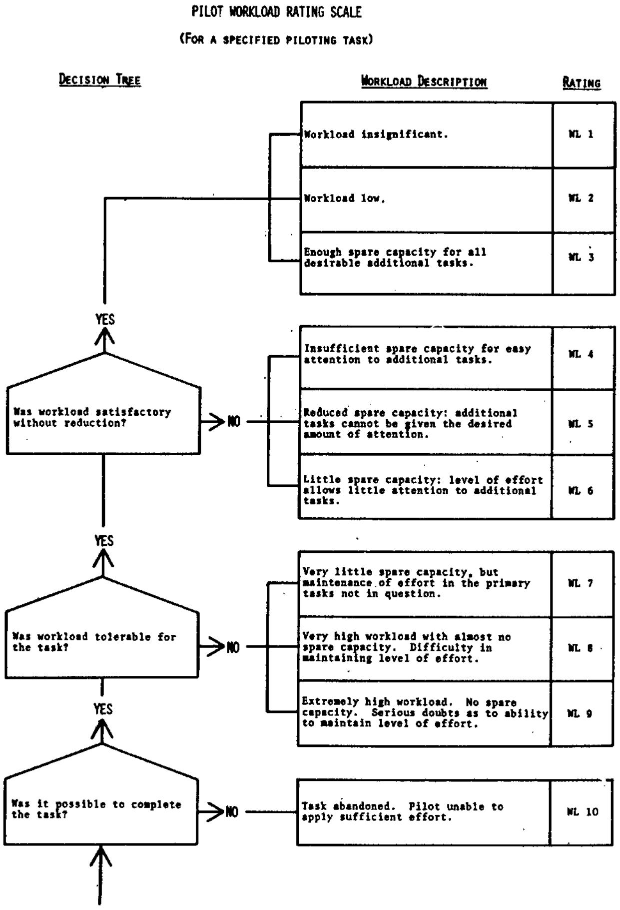


Fig 8 One pilot's workload ratings and heart rate levels for manual landings in fog



WORKLOAD DESCRIPTION	RATING
Workload insignificant.	WL 1
Workload low.	WL 2
Enough spare capacity for all desirable additional tasks.	WL 3
Insufficient spare capacity for easy attention to additional tasks.	WL 4
Reduced spare capacity: additional tasks cannot be given the desired amount of attention.	WL 5
Little spare capacity: level of effort allows little attention to additional tasks.	WL 6
Very little spare capacity, but maintenance of effort in the primary tasks not in question.	WL 7
Very high workload with almost no spare capacity. Difficulty in maintaining level of effort.	WL 8
Extremely high workload. No spare capacity. Serious doubts as to ability to maintain level of effort.	WL 9
Task abandoned. Pilot unable to apply sufficient effort.	WL 10

Fig 9 Pilot workload rating scale

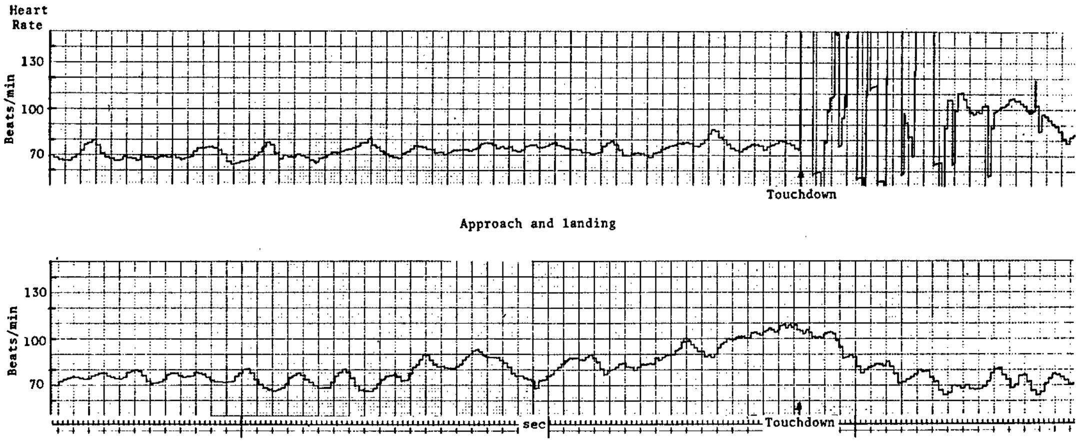


Fig 10 Heart rates recorded during a sortie of experimental 6^o approaches using DLC (BAC 1-11)

Upper trace - Approach ending in a heavy landing
Lower trace - Typical response for approach and 'roller' landing

SOME PHYSIOLOGICAL CHANGES IN LATERAL (Gy) ACCELERATION

Major James R. Popplow, M.D.

Air Force Aerospace Medical Research Laboratory
Aerospace Medical Division, Air Force Systems Command
Wright-Patterson Air Force Base, Ohio

ADDRESS FOR CORRESPONDENCE: AFAMRL/BBS
Wright-Patterson AFB, OH 45433

ABSTRACT

The advent of the Advance Fighter Technology Integration (AFTI) F-16 has prompted a renewed interest in the physical and physiological stresses which occur in lateral acceleration. The Aerospace Medical Research Laboratories' Acceleration Effects Branch has recently completed this experimental protocol which examined the cardiopulmonary effects of Gy forces.

Eight human male subjects, who were all permanent members of the AFAMRL centrifuge subject volunteer group, were used as subjects. They were exposed to lateral acceleration forces of 1-1/2, 2 and 2-1/2 Gy, both left and right, in combination with either 1 or 2 Gz. Recorded data included; peripheral arterial oxygen saturation, heart rate, respiratory rate and profile, tracking task error and shoulder pad pressures. This presentation will discuss the physiological parameters.

Our preliminary analysis indicates a marked reduction in pulmonary function which is linearly related to the level of lateral force. Using only these modest levels of Gy, a progressive reduction in peripheral oxygen saturation and a concomitant increase in heart rate and respiratory rate were observed.

These results are contrasted with the equivalent levels of vertical (Gz) and transverse (Gx) accelerative forces and a discussion of the possible adverse effects on the pilot are summarized.

PARAMETERS MEASURED:

- ARTERIAL OXYGEN SATURATION
- HEART RATE
- RESPIRATORY RATE
- INSPIRATORY VOLUME
- INSPIRATORY TIME

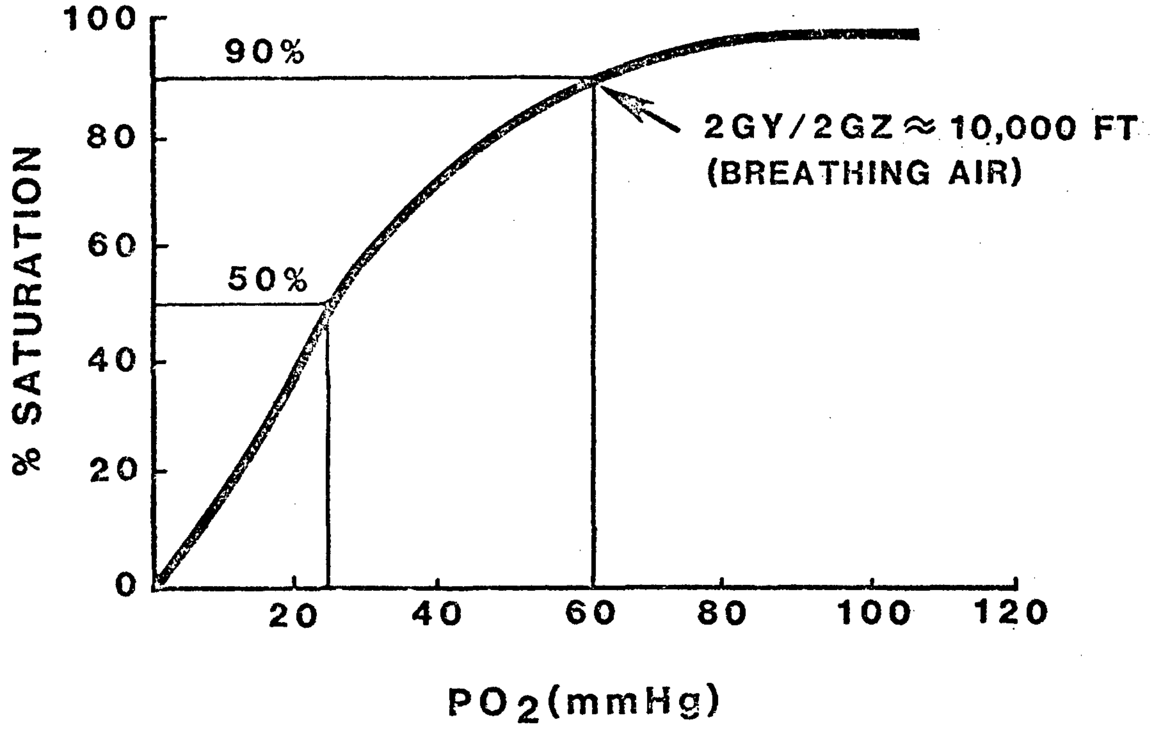
ADDITIONAL STUDIES:

TRACKING PERFORMANCE

SHOULDER PAD PRESSURES

SUBJECTIVE QUESTIONNAIRE

Hemoglobin Oxygen Dissociation Curve



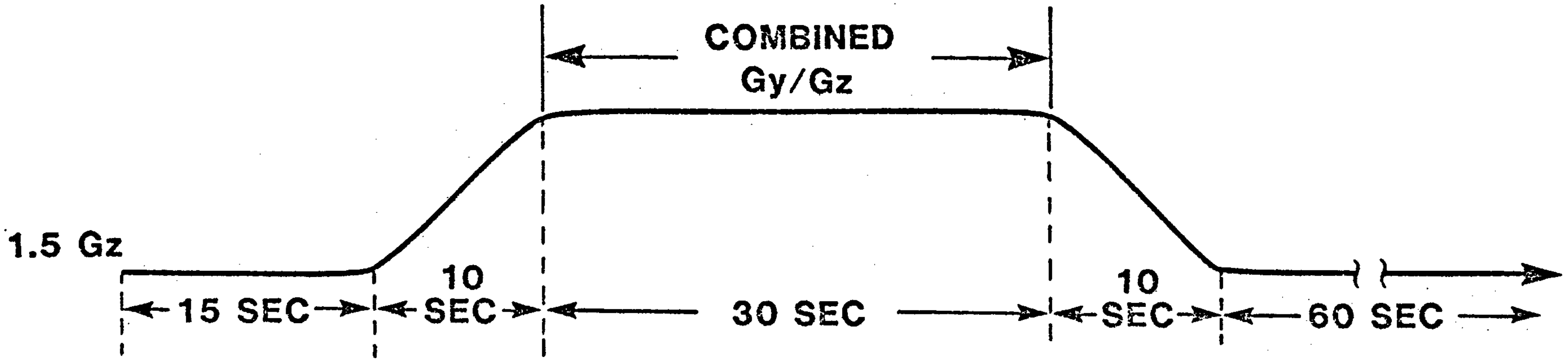
PO_2	% SATURATION
100	97.4
90	96.9
80	95.9
70	94.1
60	90.9
50	85.1
40	74.7
30	57.5
20	32.4
10	9.6

SaO_2 and Pulmonary Shunt Values in Acceleration

	Breathing Air		Breathing 99.6% O_2			
	Aortic SaO_2	Dep Pul Vein SaO_2	Aortic SaO_2	Dep Pul Vein SaO_2	Aortic Shunt	Dep Pul Vein Shunt
+6 GX	71	61	97	93	36	46
-6 GX	74		98		41	
+5 GY	72	55	76	63	65	90
-5 GY	85	63	80	67	61	93
+7 GY	57	46	72	53	65	98
-7 GY	75	54	78	68	62	88

From Vandenberg et al., 1968

Table 20



PARAMETER READINGS EVERY FIVE SECONDS AVERAGED OVER THE FIVE SECOND INTERVAL.

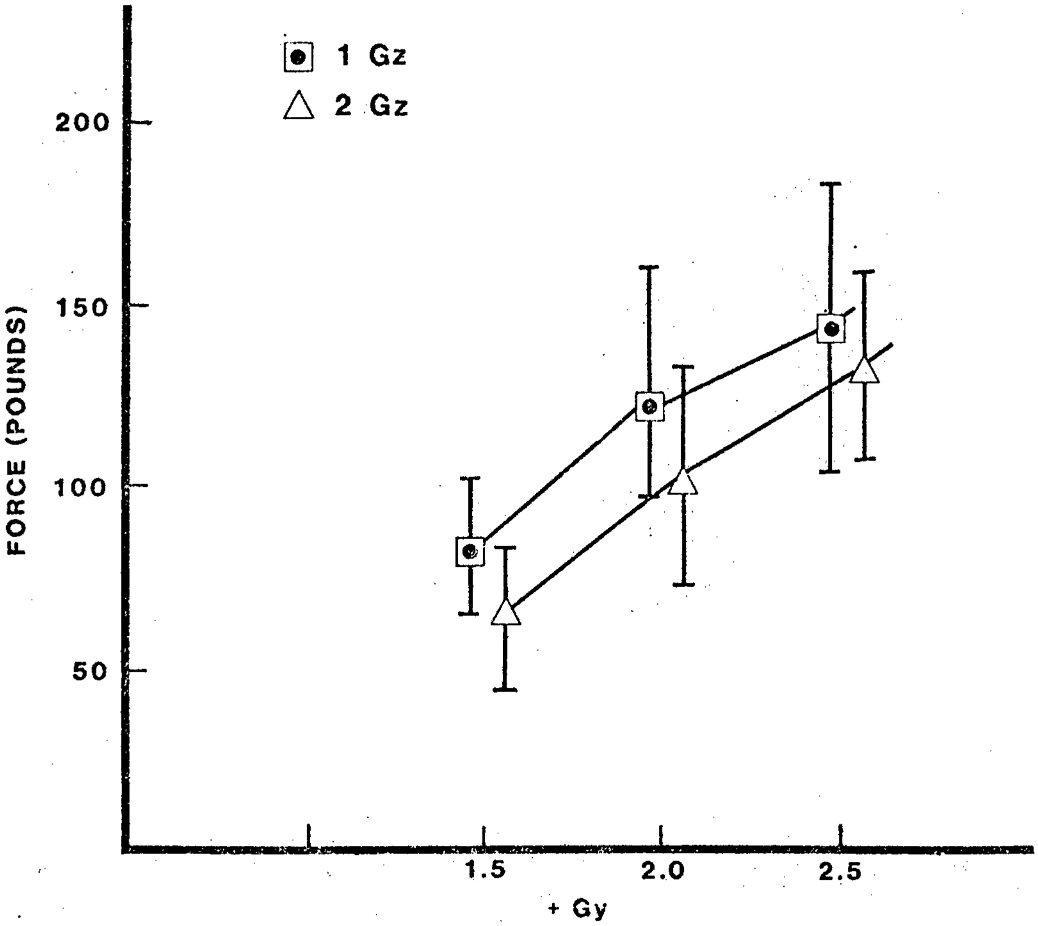
Figure 1. *Acceleration Profile*

Gy/Gz COMBINATIONS:

+ 1.5 Gy WITH 1 Gz AND 2 Gz
+ 2.0 Gy WITH 1 Gz AND 2 Gz
+ 2.5 Gy WITH 1 Gz AND 2 Gz

- 1.5 GY WITH 1 Gz AND 2 Gz
- 2.0 Gy WITH 1 Gz AND 2 Gz
- 2.5 Gy WITH 1 Gz AND 2 Gz

EIGHT RANDOMIZED RUNS PER DAY X 3 DAYS



LEFT SHOULDER PAD FORCES MEASURED DURING THE LATERAL G CENTRIFUGE STUDIES. MEANS AND STANDARD DEVIATION FOR 8 SUBJECTS.

+Gy	+Gz	POUNDS	S.D.
1.0	0	63	±20
1.5	1	83	±19
1.5	2	67	±20
2.0	1	123	±28
2.0	2	104	±32
2.5	1	146	±40
2.5	2	134	±26

TABLE 1. Left shoulder pad forces measured during lateral G centrifuge studies. Means and standard deviation. N=8.

SUBJECTIVE QUESTIONNAIRE

	NONE	MIN	MOD
NECK FATIGUE			
BREATHING DIFFICULTY			
O_2 MASK DIFFICULTIES			
DIZZINESS/DISORIENTATION			
TRACKING DIFFICULTY			
BODY MOVEMENT			

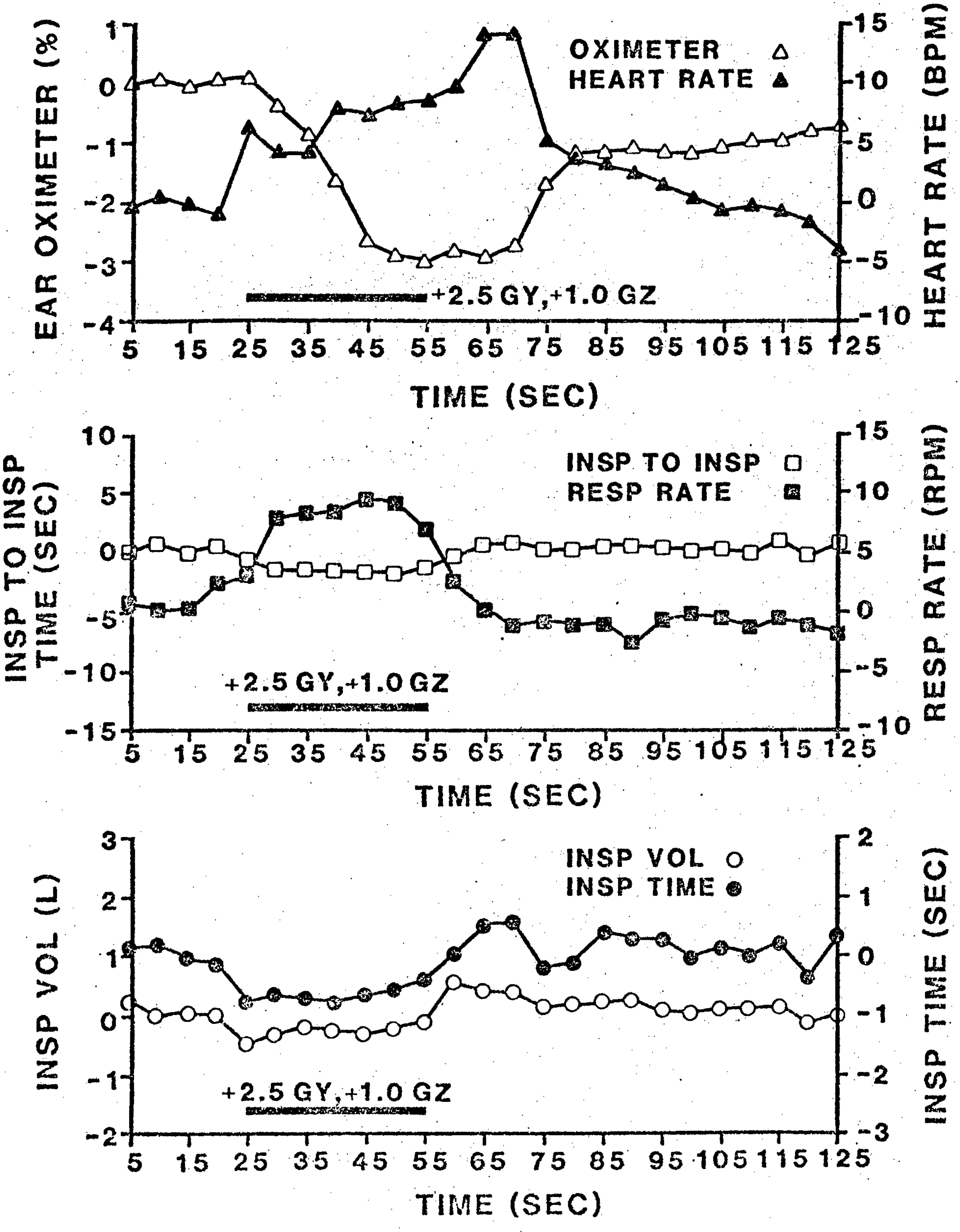

EAR OXIMETER (%)
HEART RATE (BPM)
OXIMETER △
HEART RATE ▲
+2.5 GY,+1.0 GZ
1
0
-1
-2
-3
-4
15
10
5
0
-5
-10
5 15 25 35 45 55 65 75 85 95 105 115 125
TIME (SEC)
INSP TO INSP TIME (SEC)
RESP RATE (RPM)
INSP TO INSP □
RESP RATE ▨
+2.5 GY,+1.0 GZ
10
5
0
-5
-10
-15
15
10
5
0
-5
-10
5 15 25 35 45 55 65 75 85 95 105 115 125
TIME (SEC)
INSP VOL (L)
INSP TIME (SEC)
INSP VOL ○
INSP TIME ●
+2.5 GY,+1.0 GZ
3
2
1
0
-1
-2
2
1
0
-1
-2
-3
5 15 25 35 45 55 65 75 85 95 105 115 125
TIME (SEC)

Figure 5

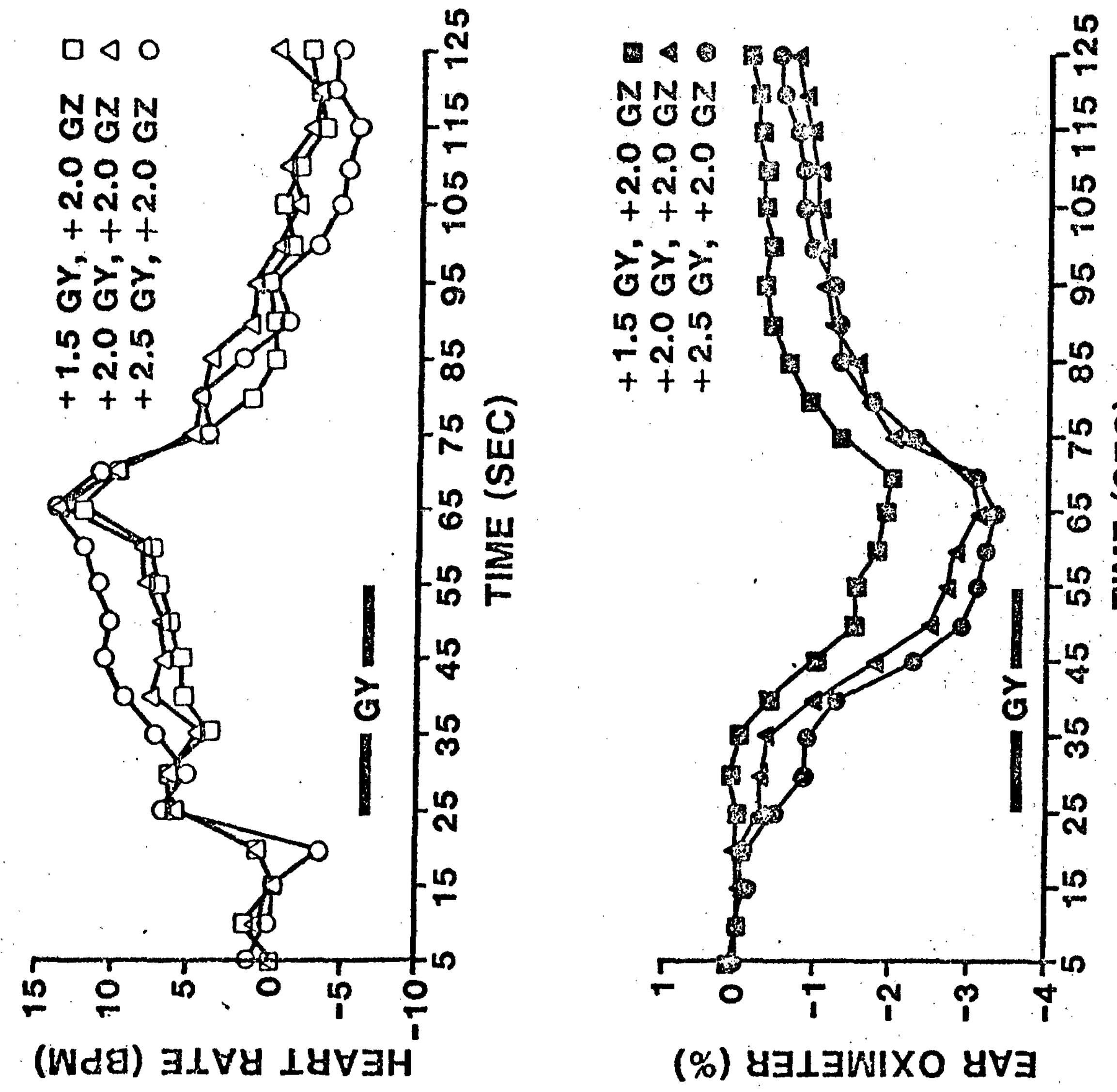

HEART RATE (BPM)
15
10
5
0
-5
-10
TIME (SEC)
5
15
25
35
45
55
65
75
85
95
105
115
125
GY
+1.5 GY, +2.0 GZ □
+2.0 GY, +2.0 GZ △
+2.5 GY, +2.0 GZ ○
EAR OXIMETER (%)
1
0
-1
-2
-3
-4
TIME (SEC)
5
15
25
35
45
55
65
75
85
95
105
115
125
GY
+1.5 GY, +2.0 GZ ■
+2.0 GY, +2.0 GZ ▲
+2.5 GY, +2.0 GZ ●

Figure 16

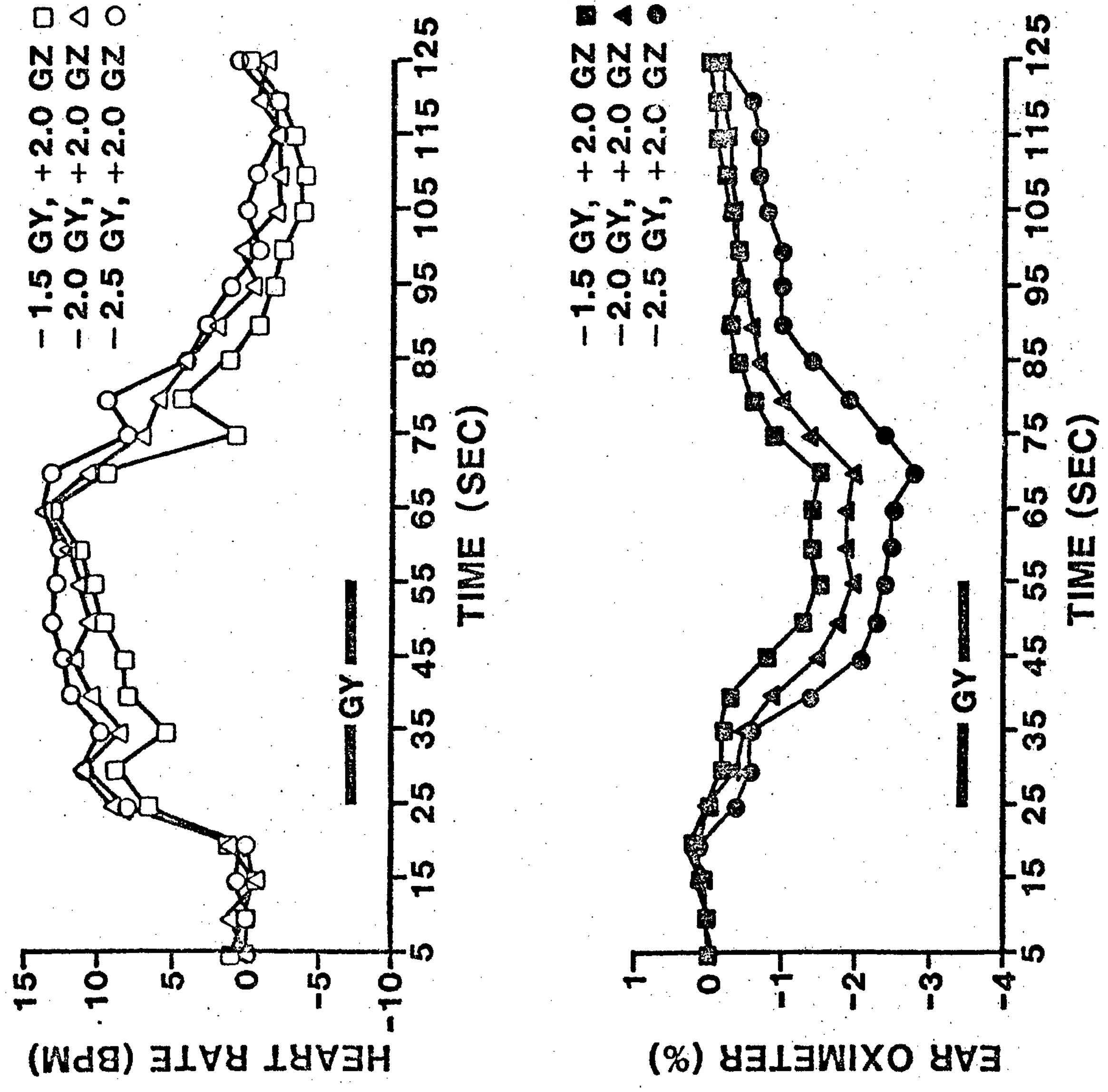

−1.5 GY, +2.0 GZ □
−2.0 GY, +2.0 GZ △
−2.5 GY, +2.0 GZ ○
HEART RATE (BPM)
15
10
5
0
-5
-10
GY
5
15
25
35
45
55
65
75
85
95
105
115
125
TIME (SEC)
−1.5 GY, +2.0 GZ ■
−2.0 GY, +2.0 GZ ▲
−2.5 GY, +2.0 GZ ●
EAR OXIMETER (%)
1
0
-1
-2
-3
-4
GY
5
15
25
35
45
55
65
75
85
95
105
115
125
TIME (SEC)

Figure 18

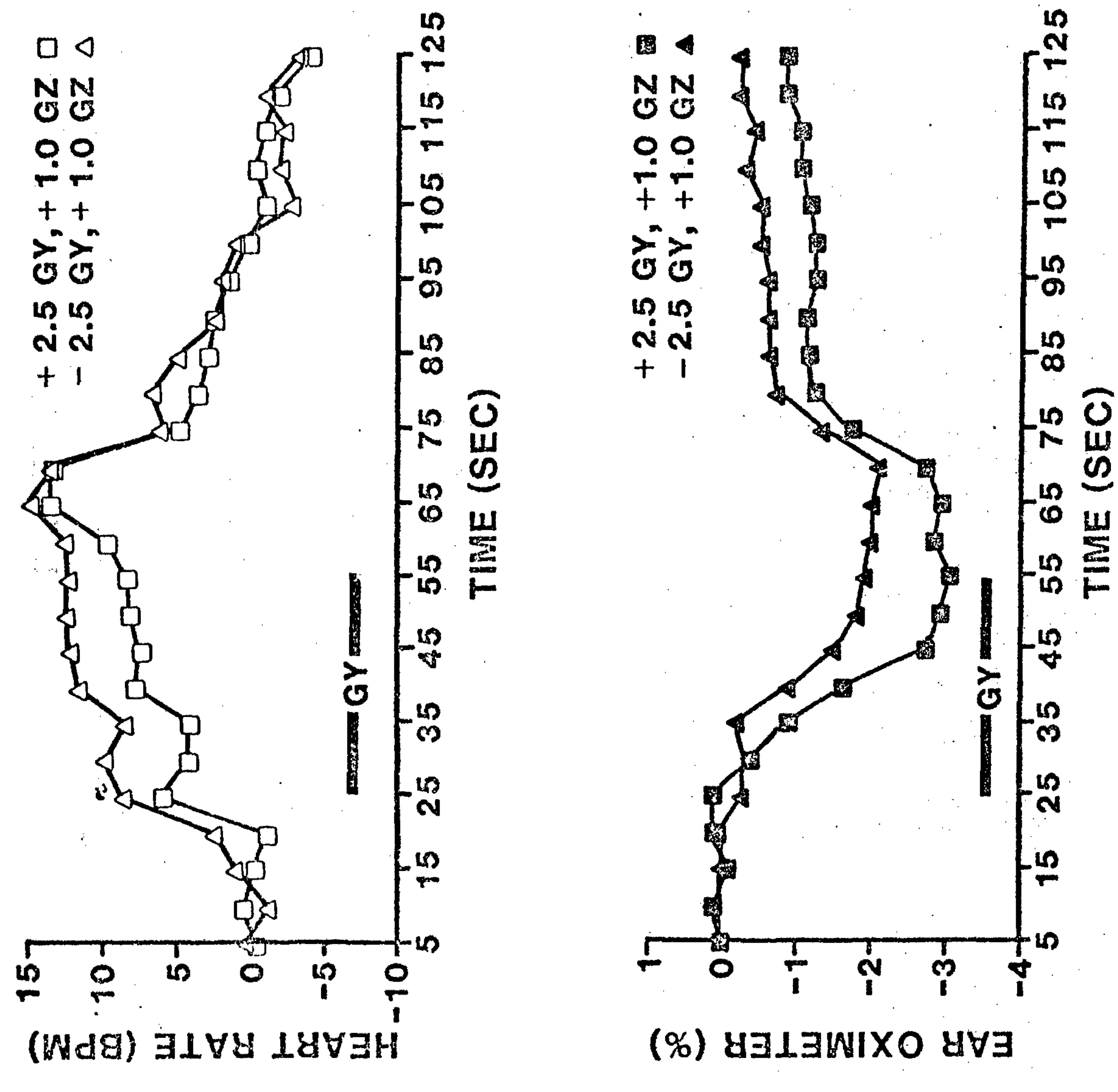

HEART RATE (BPM)
15
10
5
0
-5
-10
5
15
25
35
45
55
65
75
85
95
105
115
125
TIME (SEC)
GY
+ 2.5 GY, + 1.0 GZ □
− 2.5 GY, + 1.0 GZ △
EAR OXIMETER (%)
1
0
-1
-2
-3
-4
5
15
25
35
45
55
65
75
85
95
105
115
125
TIME (SEC)
GY
+ 2.5 GY, +1.0 GZ ▨
−2.5 GY, +1.0 GZ ▲

Figure 21

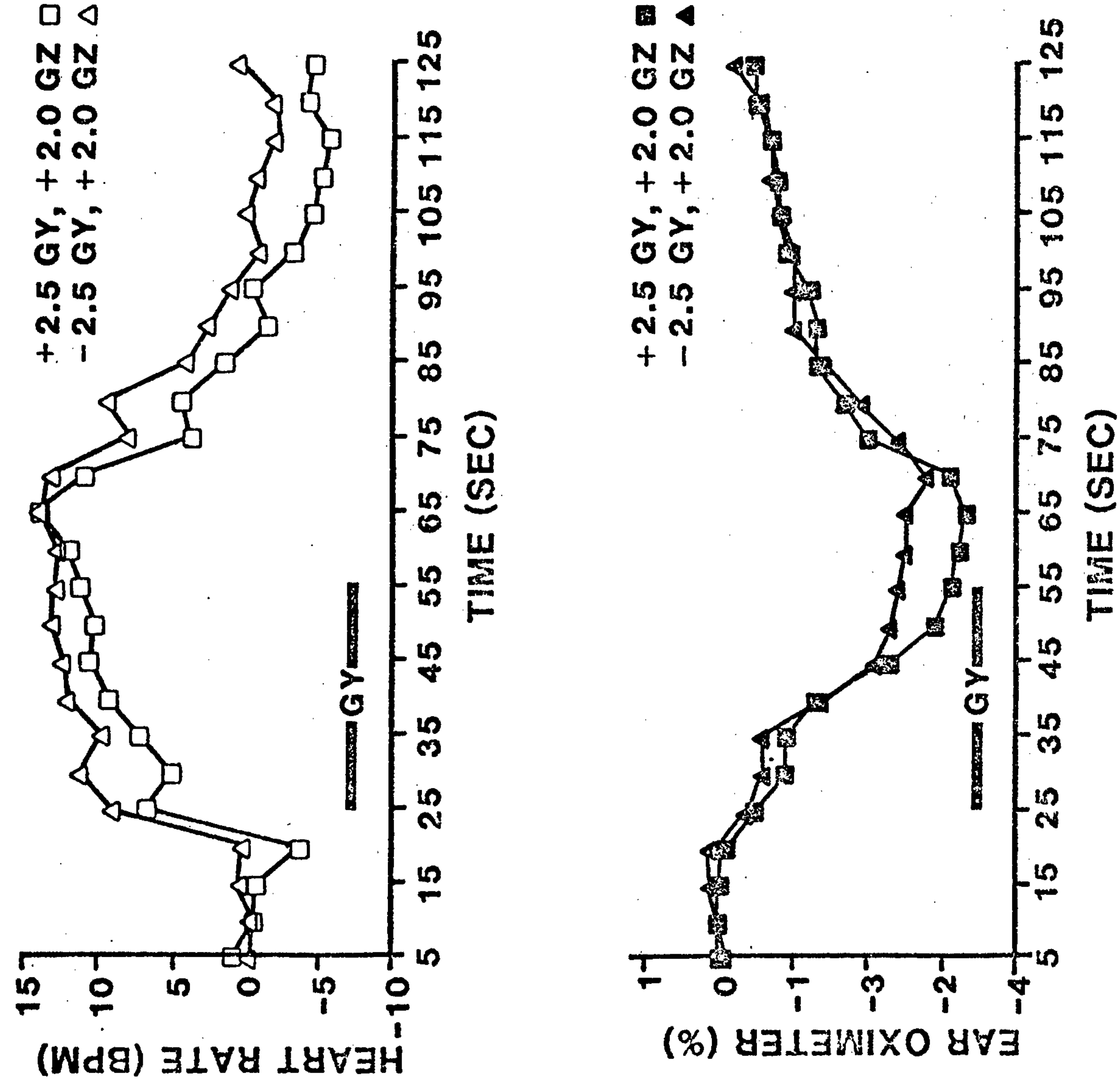

HEART RATE (BPM)
15
10
5
0
-5
-10
5
15
25
35
45
55
65
75
85
95
105
115
125
TIME (SEC)
GY
+2.5 GY, +2.0 GZ □
−2.5 GY, +2.0 GZ △
EAR OXIMETER (%)
1
0
-1
-3
-2
-4
5
15
25
35
45
55
65
75
85
95
105
115
125
TIME (SEC)
GY
+2.5 GY, +2.0 GZ ■
−2.5 GY, +2.0 GZ ▲

Figure 24

Heart Rate at Peak GY/2GZ

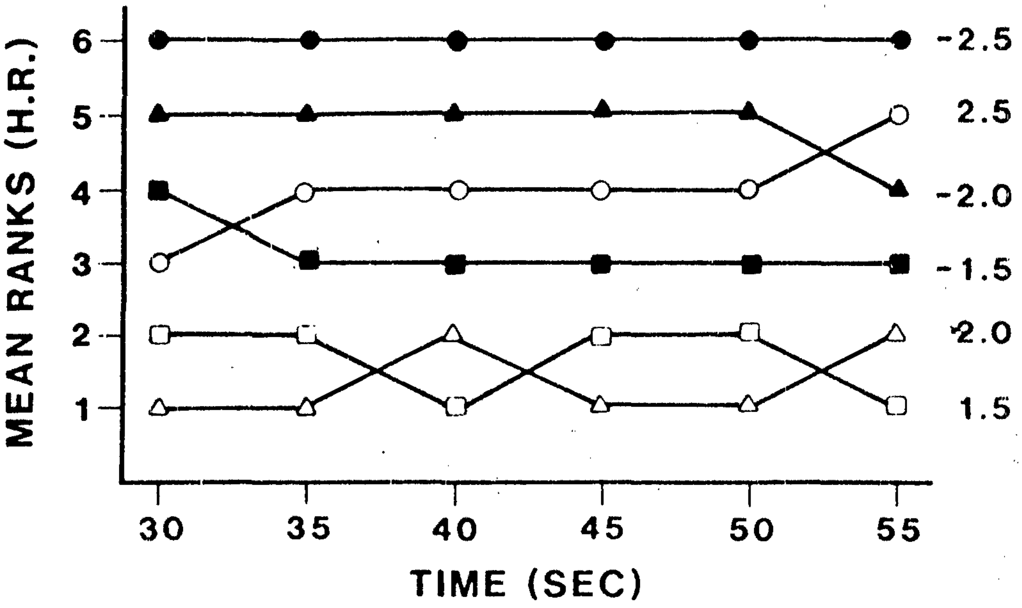


Table 5

DUNCAN'S MULTIPLE RANGE TEST*(ALPHA=0.05)

Mean Heart Rate at Peak GY/2GZ

TIME (SEC)							
	GY	2.0	1.5	2.5	-1.5	-2.0	-2.5
30	MEAN	81.5	82.38	85.73	87.75	92.07	95.19
	N	16	16	15	16	15	16
	GY	2.0	1.5	-1.5	2.5	-2.0	-2.5
35	MEAN	79.63	80.00	82.81	86.25	89.25	89.63
	N	16	16	16	16	16	16
	GY	1.5	2.0	-1.5	2.5	-2.0	-2.5
40	MEAN	82.06	83.81	86.06	89.50	90.81	94.13
	N	16	16	16	16	16	16
	GY	2.0	1.5	-1.5	2.5	-2.0	-2.5
45	MEAN	82.56	82.8	86.56	91.94	93.56	95.75
	N	16	15	16	16	16	16
	GY	2.0	1.5	-1.5	2.5	-2.0	-2.5
50	MEAN	83.38	83.50	88.31	92.19	92.88	98.03
	N	16	16	16	16	16	16
	GY	1.5	2.0	-1.5	-2.0	2.5	-2.5
55	MEAN	84.44	85.56	89.19	94.50	94.81	101.06
	N	16	16	16	16	16	16

*NOTE:
Items connected by underlining are not significantly different.

Table 6

SaO_2 at Peak GY/2GZ

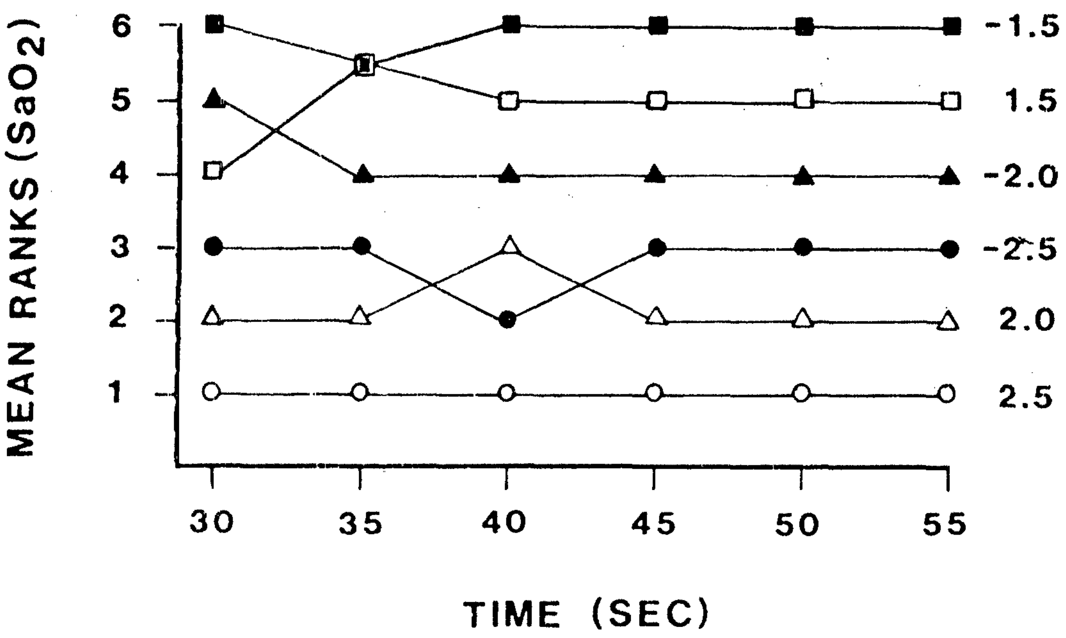


Table 14

DUNCAN'S MULTIPLE RANGE TEST (ALPHA= 0.05)

Mean SaO_2 at Peak GY/2GZ

TIME (SEC)							
	GY	2.5	2.0	-2.5	-2.0	-1.5	1.5
30*	MEAN	93.48	93.95	94.05	94.37	94.59	94.72
	N	15	16	16	15	16	16
	GY	2.5	2.0	-2.5	-2.0	-1.5	1.5
35*	MEAN	93.43	93.84	93.98	94.33	94.66	94.66
	N	16	16	16	16	16	16
	GY	2.5	-2.5	2.0	-2.0	1.5	-1.5
40	MEAN	92.87	93.19	93.21	93.94	94.28	94.46
	N	16	16	16	16	16	16
	GY	2.5	2.0	-2.5	-2.0	1.5	-1.5
45	MEAN	91.58	92.23	92.29	93.24	93.7	93.97
	N	16	16	16	16	15	16
	GY	2.5	2.0	-2.5	-2.0	1.5	-1.5
50	MEAN	90.63	91.35	91.75	92.90	93.20	93.53
	N	16	16	16	16	16	16
	GY	2.5	2.0	-2.5	-2.0	1.5	-1.5
55	MEAN	90.11	90.92	91.46	92.74	93.16	93.29
	N	16	16	16	16	16	16

* NOTE:
Items connected by underlining are not significantly different.

Table 15

SaO_2 *Values in Lateral Acceleration*

G	Time	Inspired Gas	Minimum SaO_2*	S.D.
−1.5 GY 1 G	30 SEC	AIR	93.9%	1.0
−1.5 GY 2 GZ	30 SEC	AIR	93.5%	1.4
+1.5 GY 1 GZ	30 SEC	AIR	93.6%	1.2
+1.5 GY 2 GZ	30 SEC	AIR	93.0%	2.0
−2.0 GY 1 GZ	30 SEC	AIR	93.5%	1.5
−2.0 GY 2 GZ	30 SEC	AIR	93.0%	1.1
+2.0 GY 1 GZ	30 SEC	AIR	92.5%	1.5
+2.0 GY 2 GZ	30 SEC	AIR	91.9%	1.2
−2.5 GY 1 GZ	30 SEC	AIR	92.9%	1.7
−2.5 GY 2 GZ	30 SEC	AIR	92.2%	1.2
+2.5 GY 1 GZ	30 SEC	AIR	92.0%	1.2
+2.5 GY 2 GZ	30 SEC	AIR	91.7%	1.2

* USING 95% AS ZERO REF.

Table 22

THE MAXIMUM EFFECT OF ACCELERATION ON SaO_2 VARIES WITH:

- ONSET RATE
- MAXIMUM G
- TIME AT MAXIMUM G

COMPARABLE ACCELERATION VECTORS

AT PEAK G FOR 30 SEC

7-10 GX APPROX 5-6 GZ APPROX 2.5 GY/2GZ

PROBING THE COGNITIVE INFRASTRUCTURE WITH EVENT-RELATED BRAIN POTENTIALS

By

Emanuel Donchin, Arthur Kramer, and Christopher Wickens

Cognitive Psychophysiology Laboratory
Department of Psychology
University of Illinois
Champaign, Illinois 61820

Paper Presented at

AIAA Workshop on
Flight Testing to Identify Pilot
Workload and Pilot Dynamics
Edwards Air Force Base, California
January 19-21, 1982

INTRODUCTION

We review in this chapter evidence that suggests that the Event-Related Brain Potential (ERP) can be incorporated in the collection of tools of Human Engineering. The utility of the ERP as a tool in the study of Cognitive Science has been discussed elsewhere (Donchin, 1975, 1979, 1981; Wickens, 1979). As the Human Factors that must be addressed by the engineer are increasingly "cognitive" in nature (Rasmussen, 1981; Sheridan, 1981), there is an increasing need for enriching the repertoire of techniques for the assessment of cognitive function. We believe that psychophysiological techniques, in particular ERP-based procedures, can serve this function. We realize that this proposition is not self-evident to the Human Engineering profession. The recording of the ERP is cumbersome. Electrodes must be placed on the subject's scalp. Special equipment is needed for analyzing, digitizing, averaging and displaying the data. The physiological nature of the signals is essentially unknown and the functional significance of the ERP components is a subject of controversy. What benefits would accrue to the system designer as she encumbers herself with this exotic technique? Is it likely to help in the assessment of workload? After all, there is a strong tendency to trust the subjective reports of operators in assessing workload. These reports appear to be preferred even to the seemingly simpler techniques proposed by the Experimental Psychologist. Sheridan and his co-workers concluded (Sheridan and Simpson, 1979; Sheridan, 1980) that it is possible to obtain a reliable and valid measure of workload by administering a rather simple questionnaire. Why should one bother with more costly, elaborate and indirect measurements of workload?

The question is reasonable and the answer is clear. If you gain nothing by complicating the measurement process it is best to avoid the complications. We claim, however, that there are circumstances in which subjective reports may need augmentation, and in a subset of these circumstances, the ERPs may be very useful.

Consider, for example, the following task. In Figure 1 are displayed four pairs of words. Your task is to write "yes" next to the pair if the words rhyme, and to write "no" next to the pair if the words do not rhyme. Please do so as rapidly as you can. Did you find it equally easy to decide in all four cases? Most subjects report that the decision requires the same effort regardless of the pair we used and are quite surprised when they find that their subjective assessment of the workload imposed by these simple judgments does not reflect objective measures of performance.

You will note that the four word pairs in Figure 1 are instances of four possible relationships between the two words in the pair, as follows:

1. (RO) The two words rhymed and looked alike (MATCH-CATCH).
2. (R-) The two words rhymed but did not look alike (MAKE-ACHE).
3. (WO) The two words looked alike but did not rhyme (CATCH-WATCH).
4. (W-) The two words neither rhymed, nor looked alike (SHIRT-WITCH).

We label these pairs with an R to indicate a phonological match, with an O to indicate an orthographic match, and W indicates a phonological mismatch and the dash (-) an

Match — Catch

Make — Ache

Catch — Watch

Shirt — Witch

Figure 1. A sample of word pairs presented to subjects in a phonological judgement task.

orthographic mismatch. Thus for the RO and W- the phonological and the orthographic information agree and for the R- and WO pairs there is a conflict between the phonological and the orthographic information. While it is easy to analyze the stimuli in Figure 1 and see that they do indeed differ in these attributes, subjects do not usually perceive themselves as having greater difficulty in deciding that the words "CATCH-WATCH" do not rhyme, than they do in deciding that the pair "SHIRT-WITCH" do not rhyme.

But these subjective impressions are somewhat misleading. Polich, McCarthy, Wang and Donchin (in preparation) and Kramer and Donchin (1982), presented subjects with the two words of the pair in succession and required them to indicate their judgments by pressing one of two buttons immediately after the appearance of the second word. The reaction times belie the subjective reports. This can be seen from Figure 2 where the reaction times for each of the classes is shown. It is clear that a conflict between the phonology and the orthography retards the subject's reaction by a considerable number of milliseconds. The delay is about 300 msec when the second word "looks like" the first word but does not rhyme with it (the WO pairs like CATCH-WATCH). In other words, an individual's subjective assessment may not reveal a processing delay that may cost an operator up to three tenths of a second in responding to a display change! Clearly, a non-trivial delay in high-performance aircraft.

What we find then is that when tasks place demands on the human information processing system that affect, or depend on, interactions

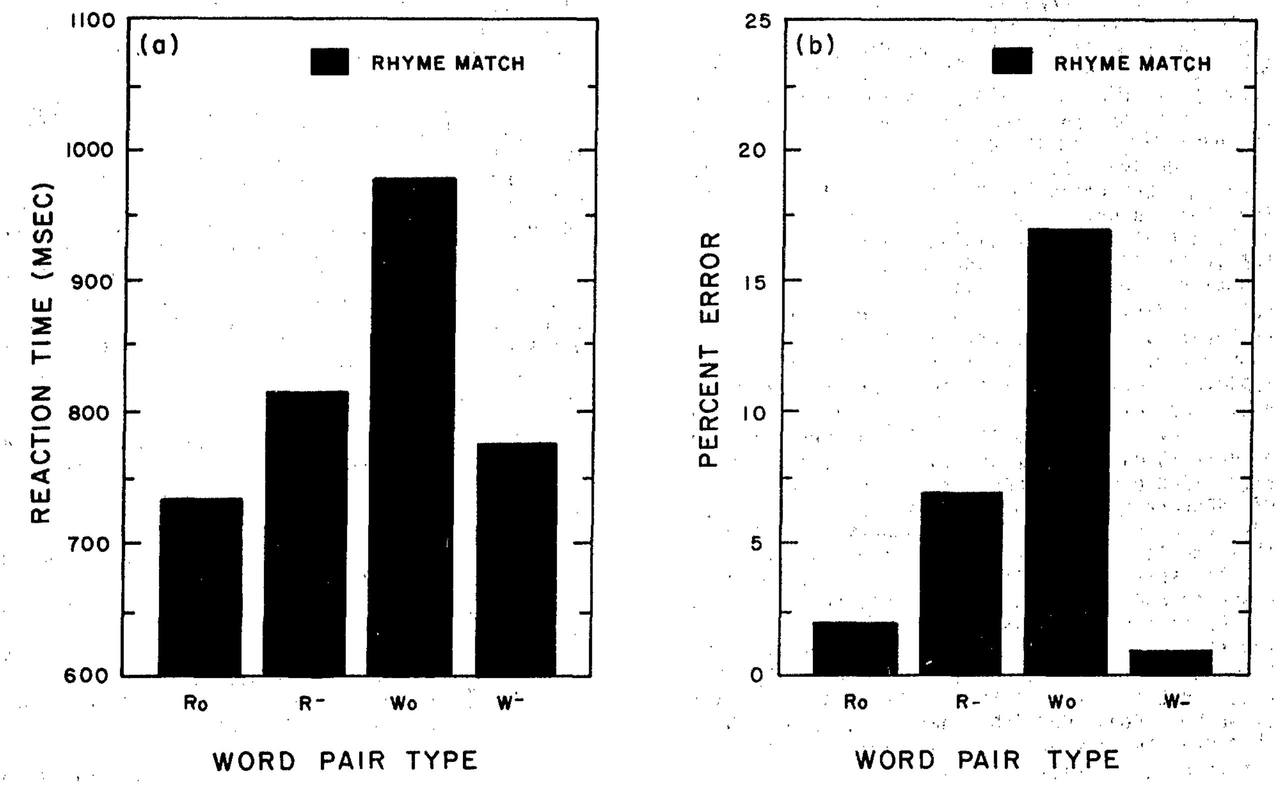


Figure 2. Mean Reaction Times for correct responses and percentage of errors averaged across 40 subjects in the phonological judgement task (After Kramer, et al., 1982).

between the automatically activated elements of the processing machinery, loads may be imposed on the system that directly affect its performance even though they are not available to the internal monitors that yield subjective reports.

This phase of the analysis illustrates the need to supplement subjective reports by accurate and detailed measures of performance. Where, though, can the Psychophysiology help? We submit that its most effective role is, when properly used, in carrying the analysis beyond the limits imposed by the examination of the more overt responses. Thus, for example, the data in Figure 2 indicate that the phonology-orthography conflict delays the reaction. But these data do not permit unequivocal conclusions regarding the functional locus of the delay. Does the conflict cause reprocessing of the signal? Are the subjects more cautious when they detect the conflict, or do they require more time to encode the stimuli? Why is the cost of conflict lower for the R- pair than it is for the WO pair? These, and similar questions, are important not merely for their theoretical significance but also because our understanding of the nature of the interference is necessary if we are to develop systematic guidelines for improving the design of displays and related systems. The analysis supported by the ERPs may be especially helpful when there is a need to resolve conflicting theories. There are those who suggest that phonologic and orthographic codes interact at the encoding stage (Meyer, et al., 1975; Schulman, et al., 1978). Others have suggested that the interference occurs at a response-selection stage (Conrad, 1978). Kramer and Donchin's ERP data, shown in part in Figure 3, provide information that compliments the reaction time data.

The waveforms shown in Figure 3 are of ERPs averaged over 40 subjects. These data were recorded at the parietal electrode, and each of the lines represents an average over one of the four classes of pairs (RO, R-, WO, W-). As usual, the ERP appears as a sequence of peaks and troughs, (often referred to as "components"). It is evident that the waveforms for the four ERPs are congruent until the presentation of the second stimulus. The subject, of course, does not know which of the four pair-classes will be used on any trial until the appearance of the second stimulus. Once this happens the waveforms diverge. It is quite evident that the ERP elicited by the WO pair is different than the other three ERPs. It is characterized by a substantial delay in the elicitation of a large positive (down-going) component relative to the appearance of a similar component in the other three ERPs. In our terminology, the latency of this peak, labeled the "P300" for reasons that will become apparent later, is increased in the WO pairs. Thus, the ERP provides additional data on the two types of orthography-phonology conflict that occur in this experiment. If we could interpret these ERP changes we may be able to gain a better understanding of the process. In fact, since the latency of the P300 component of the ERP provides a measure of mental processing time that is unaffected by response selection and execution processes (McCarthy and Donchin, 1981) the data of Figure 3 suggest at least some of the effect of orthography-phonology conflict operates prior to the response selection stage. It is interesting to note that the differences in the reaction time are larger than the P300 difference. This suggests that interference which begins prior to the response selection stage is amplified during later processing and therefore may reflect a cascading process (McClelland, 1979).

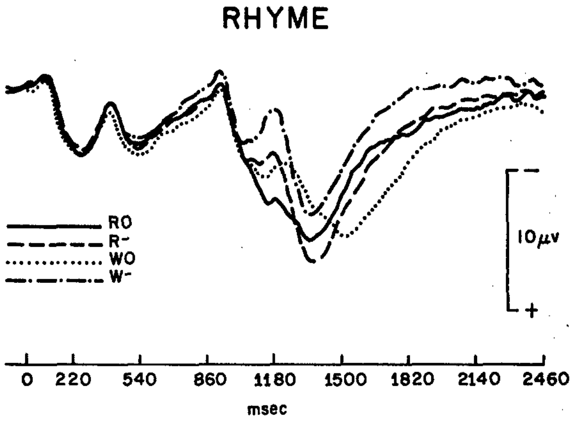


Figure 3. Average ERPs recorded at Pz. The ERPs were averaged across 40 subjects. Data span an epoch that begins 100 msec prior to the presentation of the first of a pair of words in a phonological judgement task and ending 2,560 msec after the second stimulus (After Kramer, et al., 1982).

Further examination of Figure 3 reveals that the four ERPs differ also in the disposition of a negative (upwards-going) peak that just precedes the P300. This peak is labeled "N200". The differences in the amplitude of the N200 component may serve to clarify some issues concerning the detection of orthographic and phonological mismatches. As can be seen in Figure 3, the largest N200s are elicited by the W- pairs, in which both the orthography and the phonology of the pair members mismatch. The R-list (phonological match, orthographic mismatch) also elicits a relatively large N200. Thus the R- and W- list which both orthographically mismatch, elicit an N200. This suggests that the detection of an orthographic mismatch may occur automatically. In fact, in an experimental condition, not shown here, the subjects were instructed to report "yes" if the words match visually regardless of the phonology. The N200 elicited in that condition by R- and the W- was identical to that elicited during the rhyme condition. On the other hand, the WO list (orthographic match, phonological mismatch) elicits an N200 only when the subject is instructed to detect rhymes. This suggests that a phonological mismatch may be detected only when the phonology of the task is relevant. In other words, the ERP data indicate that a phonological comparator is involved solely in the rhyme condition, even though orthographic comparators are involved regardless of the task. Thus both the latency of the P300 component and the amplitude of the N200 component provide information which compliments introspection and traditional overt response analysis. The ERP data should not be viewed as supplanting the information garnered from traditional measures but rather as a source of complimentary information.

We discussed these data because they illustrate our basic contention. Subjective reports, while valuable, do have limitations. In assessing the demands that a system places on an operator it is particularly unwise to trust introspective claims that <u>deny</u> differences in workload between the systems under comparison. This is especially so when the demands imposed by the system operate at levels of processing that are not normally open to examination by introspection. It is in this domain that the Human Factors expert is most likely to benefit from the models and techniques of the Experimental Psychologist. On occasion it will be found that the assessment can be augmented by utilizing ERPs. This is particularly true when there is an interest in developing a theoretical account for the differences between the demands imposed on the operator by different systems. The theoretical models that can be adduced abound in references to internal processing entities. As the ERP components are manifestations of such processing entities their study is of use.

In the remainder of this chapter we shall illustrate these concepts by reviewing a series of studies demonstrating that the amplitude of the P300 can serve as a measure of "workload". We shall precede this discussion with a brief overview of the study of ERPs. For more details the reader is referred to Callaway, Tueting and Koslow (1978), Otto (1979), and Donchin (in press).

<u>Introductory Comments on the P300 Component</u>

The ERP is a transient series of voltage oscillations in the brain that can be recorded from the scalp in response to the occurrence of a discrete event (Donchin, 1975; Regan, 1972). The ERP is viewed as a sequence of components commonly labeled with an "N" or a "P" denoting polarity, and a number which indicates their minimal latency measured from the onset of the eliciting event (e.g., N100 is a negative going component which occurs at least 100 msec after a stimulus). Since ERPs are relatively small, relative to the ongoing EEG (2 - 20mv for the ERPs vs. >50mv for the EEG), the study of ERPs became practical only after the development of reliable signal averagers (Clynes and Kohn, 1959). These capitalize on the fact that the ERP is, by definition, time-locked to the eliciting event.

It is crucial to recognize the componential nature of the ERP. Early studies of the ERP which treated the waveform as a unitary entity, measuring the amplitude over the entire recording epoch (Satterfield, 1965), were difficult to interpret. The effects of the experimental manipulations tend to be quite specific to a few components and a combination of the measures of the entire epoch may obscure the relevant variance. There is a degree of controversy as to the proper identification and definition of components, (Donchin, Ritter and McCallum, 1978; Picton and Stuss, 1980). In this chapter, however, we shall follow Donchin et al's (1978) definition of an ERP component in terms of the responsiveness of the waveforms to specific experimental manipulations. A component is thus mapped into a cognitive space populated by psychological concepts such as decisions, expectations, plans, strategies, associations and memories. Specific components are associated with particular entities in this cognitive space in much the same manner in which cells in the periphery of the visual system are mapped into a field in the visual cortex. The subset of elements in cognitive space associated with a particular component thus contributes to the definition of the ERP component.

The specific attributes of a waveform that are examined in defining a "component" are the amplitude, latency, and scalp distribution. It is the sensitivity of these attributes to experimental manipulations that define an ERP component. Although no reference has been made to the underlying neural source of components, it is generally assumed that a scalp distribution which is invariant across repeated stimulus presentations implies a specific and fixed set of neural generators (Goff, Allison and Vaughan, 1978; Woods and Allison, 1981). Thus the scalp distribution which is related to the underlying neural population responsible for the generation of the component is assumed to be a crucial defining characteristic.

The ERP components we discuss in this report are "endogenous" and are distinct from another class of ERPs called "exogenous" (Sutton, et al., 1965; Donchin, et al., 1978). The Exogenous components represent an obligatory response of the brain to the presentation of a stimulus. These components are primarily sensitive to such physical attributes of the stimuli as intensity, modality and rate. The seven peaks or "bumps" which occur in the first 8-10 msec after the presentation of an auditory or somatosensory stimulus are a prototypical example of the exogenous category (Jewett, Romano and Williston, 1970).

The endogenous components, on the other hand, are sensitive to the processing demands of a task rather than the physical characteristics of the stimuli. Endogenous components, typically, are not sensitive to changes in the physical characteristics of the eliciting stimuli. On the other hand, these components are very sensitive to changes in the processing demands of the task imposed on the subject. The endogenous components are nonobligatory responses to stimuli. The strategies and expectancies of the subject as well as other psychological aspects of the task account for the variance in the endogenous components. A typical example, and one to which we shall devote the remainder of this chapter, is the P300 component.

This ERP component is elicited by rare, task relevant, stimuli. A task in which it is readily elicited is often called the "oddball" paradigm. In a study by Duncan-Johnson and Donchin (1977), using this paradigm, the subject was instructed to count covertly the total number of higher pitched tones in a Bernoulli series. In different blocks of trials the relative probability of the two tones was manipulated. It can be seen from Figure 4 that the amplitude of the P300 increases monotonically as the probability of the stimulus decreases. This occurs regardless of which of the two stimuli is being counted. The basic relationship between P300 amplitude and stimulus probability is very robust and has been replicated numerous times in several variants of the classic oddball paradigm. The only requirement is that the subject count one of the two events.

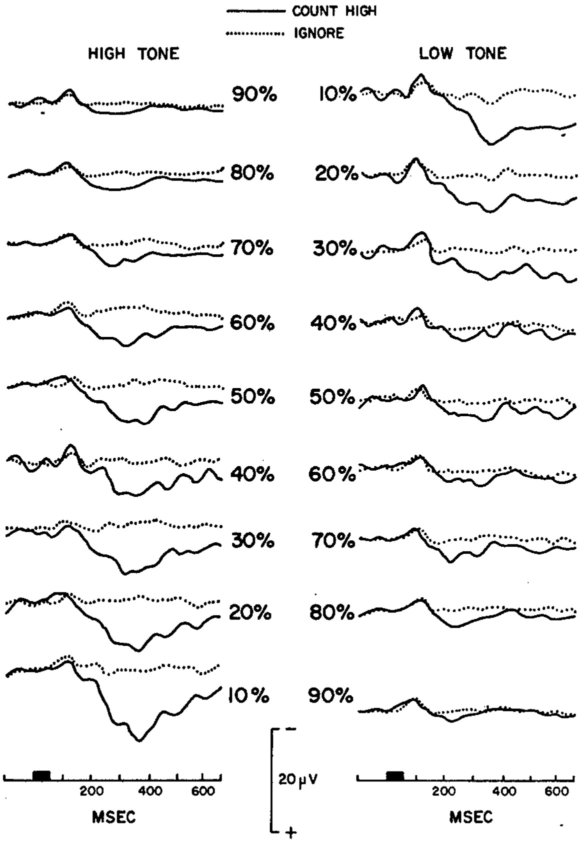


Figure 4. Averaged ERPs elicited by high and low tones presented in a Bernoulli series. The waveforms represent experimental conditions in which subjects counted the high tones (solid lines) or solved a word puzzle as the tones were presented (dotted lines) (From Duncan-Johnson and Donchin, 1977).

When the subjects were solving a word puzzle and were not required to process the tones the P300s were not elicited. Note that the ERPs in Figure 4 that were obtained in this "ignore" condition show no P300 at all levels of probability. Thus the amplitude of P300 is determined by a combination of the task relevance and the subjective probability of the eliciting event. This basic finding plays a crucial role in the use of P300 in the assessment of workload.

The demonstration that P300 is elicited by unexpected, task relevant stimuli led Donchin, McCarthy, Kutas and Ritter (in press) to suggest that "the P300 is a manifestation, at the scalp, of neural action that is invoked whenever the need arises to update the 'neuronal model' (Sokolov, 1969) that seems to underlie the ability of the nervous system to control behavior". The neural or mental model is continually assessed for deviations from inputs and revised when the discrepancies exceed some criterion value. The frequency with which the mental model is revised is based on the surprise value and task relevance of the stimuli. Donchin (1981) also argued that the concept of a subroutine is an appropriate metaphor for the activity of ERP components (Donchin, Kubovy, Kutas, Johnson and Herning, 1973; Donchin, 1975). In software applications, subroutines represent algorithms which are designed to accomplish a specific task and which can be employed in a variety of different programs. ERP components may be associated with specific information processing functions which are activated in a variety of different tasks. In the case of the P300, the subroutine may be invoked whenever there is a need to evaluate surprising, task relevant events. This interpretation of the changes in P300 amplitude is strengthened by the evidence that has accumulated in the past decade regarding the factors that control the latency of the P300. As the use we make of P300 in the analysis of man-machine interaction depends strongly on our theoretical interpretation of the component it will be useful to provide a brief review of the latency data and their interpretation.

The Latency of the P300 Component

The peak latency of the P300 component appears to depend on the time required to recognize and evaluate a task-relevant event. The latency ranges between 300 to 750 msec following the presentation of a discrete stimulus. In fairly simple tasks calling, for example, for a discrimination between two tones that differ in pitch (i.e., 1000-1600 Hz) the stimuli elicit relatively short latency P300s. More difficult discriminations (i.e., semantic analysis) result in increases in the latency of P300.

Assuming that manual or vocal reaction time terminates processing, and that P300 is a manifestation of a process that preceeds the response then it would be expected that P300 latency and reaction time should positively covary. This prediction has been supported by numerous studies (Wilkinson and Morlock, 1967; Bostock and Jarvis, 1970, Rohrbaugh, Donchin and Erikson, 1974). Other investigations, however, failed to detect a relationship between P300 latency and reaction time (Karlin, Martz and Morkoff, 1970; Karlin and Martz, 1973).

Donchin et al. (1978) proposed an interpretation of the processes underlying the P300 which may reconcile these contradictory findings. They suggested that P300 latency is determined by the time required to evaluate the stimulus, but is largely independent of response selection and execution time. The correlation between reaction time and P300 latency would, accordingly, vary as a function of the percent of reaction time variance that is accounted for by stimulus evaluation processes. This percentage would be affected by the strategies employed by the subject. The strategies, therefore, should influence the relationship between P300 latency and reaction time (see also Ritter, et al., 1972). Evidence that P300 is determined by the amount of time required to recognize and evaluate a stimulus has been reported by several investigators who employed Sternberg's (1966, 1969a, 1969b) additive factors methodology (Gomer, Spicuzza and O'Donnell, 1976; Ford, Roth, Mohs, Hopkins and Kopell, 1979; Ford, Mohs, Pfefferbaum and Kopell, 1980). Sternberg's paradigm involves the factorial manipulation of two or more experimental variables which are expected to differentially affect the durations of specific stages of processing. For example, the superimposition of a mask over a display is assumed to influence processing in an early, perceptual stage. On the other hand, reduction of the compatibility between the stimulus and the response would be expected to affect the selection and the execution of the response. In the studies mentioned above, both P300 latency and reaction time increased monotonically with increasing memory load.

Other investigators, employing different paradigms also report that P300 latency and reaction time are positively correlated when stimulus evaluation time is manipulated. N. Squires et al. (1977) found that P300 latency and reaction time covaried with the difficulty of auditory and visual discriminations. Furthermore, P300 latency varied with the manipulation of stimulus discriminability while reaction time was influenced by both stimulus evaluation and response selection factors. Heffley, Wickens and Donchin (1978) performed an experiment in which subjects were required to monitor a dynamic visual display for intensifications of one of two classes of targets. P300 latency was found to increase monotonically with the number of elements on the display. Since subjects were not required to make an overt response the differences in P300 latency were attributed to stimulus evaluation processes.

If P300 latency is determined by stimulus evaluation time and is largely independent of the time required for response selection and execution, then experimental variables which have a different effect on processing time in the two stages should influence the relationship between P300 latency and reaction time. For example, when subjects are instructed to respond quickly with a low regard for accuracy, their responses are probably emitted without full evaluation of the stimulus (Wickelgren, 1977). On the other hand, if subjects are instructed to respond accurately they are likely to perform a more thorough analysis of the stimuli prior to responding. This analysis leads to the prediction that the correlation between P300 latency and reaction time would vary with the subject's strategies. Specifically, the correlation would be high and positive when the subjects are instructed to be accurate. Low correlations would be observed under speed instructions.

Kutas, McCarthy and Donchin (1977) required subjects to distinguish between two stimuli under both speed and accuracy instructions. In one experimental condition subjects were required to discriminate between two names, Nancy and David, presented on a CRT (with relative frequencies of 20 and 80%, respectively). In a second condition female names comprised 20% of the items and males names 80%. In the third condition subjects were required to discriminate between synonyms of the word "Prod" which occurred with a relative probability of 20% and unrelated words which were presented with the complementary probability. The average P300 latency was shortest for the first condition, intermediate for the second and longest for the third condition. The more complex the discrimination, the longer the P300 latency. A detailed analysis of the single trials (Woody, 1977) revealed that the correlation between P300 latency and reaction time was larger for the accuracy condition (.617) than the speed condition (.257). Kutas et al. (1977) concluded that the data supported the hypothesis that P300 latency reflected the termination of a stimulus evaluation process while reaction time indexed the entire sequence of processing from encoding to response selection and execution. Thus, under the accuracy condition when response selection is contingent on stimulus evaluation processes P300 latency and reaction time are tightly coupled. However, when subjects perform the discrimination under the speed instructions the processes of stimulus evaluation and response selection are more loosely coupled and hence the relationship between P300 latency and reaction time is not as high.

Additional evidence bearing on the issue of the P300's sensitivity to the manipulation of stimulus evaluation processes has been obtained in a study by McCarthy and Donchin (1980) who manipulated orthogonally two independent variables in an additive factors design (Sternberg, 1969). One factor, stimulus discriminability, has been shown to affect an early encoding stage of processing while the second factor, stimulus response incompatibility, influences the later stages of response selection and execution (Bertlson, 1963; Sanders, 1970; Shwartz, Pomerantz and Egeth, 1977). The subject's task was to decide which of two target stimuli, the words RIGHT or LEFT, were presented in a matrix of characters on a CRT. The characters were either presented within a 4x4 matrix of # signs (no noise condition) or in a 4x4 matrix of letters chosen randomly from the alphabet (noise condition). Stimulus response incompatibility was manipulated by preceeding the target matrix either with the cue SAME or with the cue OPPOSITE. SAME signaled a compatible response. The cue OPPOSITE indicated an incompatible response; the right hand would respond to the word LEFT and the left hand to the cue RIGHT. Reaction time increased when the command word was embedded in noise and when the response was incompatible with the stimulus. The effect of the two variables on the RT was additive implying that these manipulations influenced different stages of processing. P300 latency was increased by the addition of the noise to the target matrix, but was not affected by the incompatibility between the stimulus and the response. These results support the conclusion that P300 latency is affected by a subset of the set of processes which affect reaction time. The P300 is elicited only after the stimulus has been evaluated. Subsequent processing required for the selection and execution of the response does not appear to influence the latency of the P300.

The P300 component of the ERP provides a metric for the decomposition of stages of information processing which compliments the traditional behavioral measures. In terms of applications to system design and workload evaluation ERPs used in conjunction with behavioral and subjective measures permit the assessment of stage specific task interference effects. For example, if two time-shared tasks interfere with each other it is usually desirable to know the locus of this interaction. Only by discovering the stage at which tasks interact can systems be designed which minimize operator workload.

THE P300 AND HUMAN ENGINEERING

P300 and Perceptual/Central Processing Resources

The studies reviewed above provided evidence that the P300 component is a manifestation, at the scalp, of a processing-entity, or a subroutine, that is involved whenever task-relevant surprising stimuli are present. The routine appears to be performing a role in the context-updating activities that occur whenever an event calls for the revision of the neuronal model or schema of the environment. It is noteworthy that this subroutine is invoked only if the stimuli are associated with a task that requires that they be processed. Ignored stimuli do not elicit a P300. But what if the stimuli are only partially ignored? What if the subject is instructed to perform the oddball task concurrently with another task? Would the amplitude of the P300 reflect the centrality of the oddball task? Would it, perhaps, change with the amount of resources allocated to the oddball task? Clearly, if so - the P300 may serve as a very useful measure of the amount of resources demanded by the two tasks. It is this series of questions that lie at the core of the usage that can be made of P300 in the assessment of workload.

The study of cognitive workload and of the allocation of processing resources to several tasks performed concurrently is, in fact, the area of research that has profited from the incorporation of ERP measures. The research reviewed here began in the Cognitive Psychophysiology Laboratory in the mid 1970's with support from DARPA. It has been performed within the framework of Resource Allocation theory. This class of models suggests that it is useful to conceptualize human capacity as represented by a finite pool of "resources" available for sharing among concurrently performed tasks (Kahneman, 1973; Moray, 1967; Norman and Bobrow, 1975). In the Kahneman (1973) model these processing resources were undifferentiated, implying that all tasks draw resources from the same pool. The general model predicts that when two tasks are timeshared their performance should decrease relative to single task levels.

This model underlies the secondary task technique, a method which is commonly employed in the assessment of the workload associated with a task as the workload is viewed as reflected by the amount of processing resources consumed by a task (Knowles, 1963; Rolfe, 1971; Wickens, 1979). In the secondary task technique, the subject is assigned two tasks; a "primary" task that is to be performed as well as possible and the second task that need be performed only to the extent that primary task performance remains stable. It is assumed that the demands imposed upon the subjects by the primary task can be assessed by monitoring performance on the secondary task. An easy primary task will require a minimal amount of processing resources, leaving an ample supply for the performance of a secondary task, while a difficult primary task will require the majority of processing resources, leaving an insufficient supply for the performance of the secondary task. Thus, the better the performance of the secondary task the less demanding the primary task.

Although the secondary task procedure has been extensively used it presents a number of practical problems (Brown, 1978; Ogden, Levine and Eisner, 1979). Particularly unfortunate is the fact that secondary task responses often intrude upon primary task performance. Of course, fluctuations in primary task performance make the interpretation of the resource trade-off extremely difficult. Evidently, it would be useful to have a secondary task which is sensitive to changes in primary task difficulty but which does not require an overt response.

It was the basic assumption of our research program that the oddball task can be used as a non-intrusive secondary task since the ERP-eliciting tones occur intermittently, are easily discriminable and do not require an overt response. Another advantage of this procedure is that it could be applied uniformly across different operational settings. In other words, the oddball task could be inserted into virtually any operational setting without requiring modifications in the system associated with the primary task. Wickens, Isreal and Donchin (1977) reported one of the first studies in the series using a compensatory tracking task as the primary task and the oddball paradigm as the secondary task.

Figure 5 illustrates the experimental procedures used in this and several other studies to be discussed. The subjects sat in front of a CRT and were instructed to cancel computer generated cursor movements by keeping the cursor superimposed on a target in the center of the display. This was accomplished by movement of a joystick mounted on the right-hand side of the subject's chair. Levels of tracking difficulty were manipulated by requiring the subject to track in either one or two dimensions (horizontal and/or vertical). The compensatory tracking task was defined as the primary task. In addition to the tracking, the subjects were also instructed to count one of two tones presented in a Bernoulli series of high and low pitched tones. Control conditions were also included in which the subjects performed each of the two tasks separately.

The data indicate that the introduction of the tracking task drastically diminishes the amplitude of the P300. However, no further reduction in P300 amplitude could be observed as tracking difficulty increased by requiring tracking in two dimensions. Even though tracking difficulty, assessed by Root Mean Square error (RMS), as well as by reaction time to the tones, definitely increased with the addition of a tracking dimension, P300 amplitude did not change. Isreal, Chesney, Wickens and Donchin (1980) conducted a similar study requiring subjects to perform a compensatory tracking task concurrently with a counting task. In this case, however, the bandwidth of the random forcing function rather than the dimensionality of the tracking task was manipulated. The bandwidth was increased gradually until the cursor's speed reached the highest level the subject could tolerate without exceeding a preset error criterion.

The results are shown in Figure 6. Again, P300 amplitude is diminished by the introduction of the tracking task, but increases in the bandwidth of the forcing function did not produce systematic changes in the amplitude of the P300.

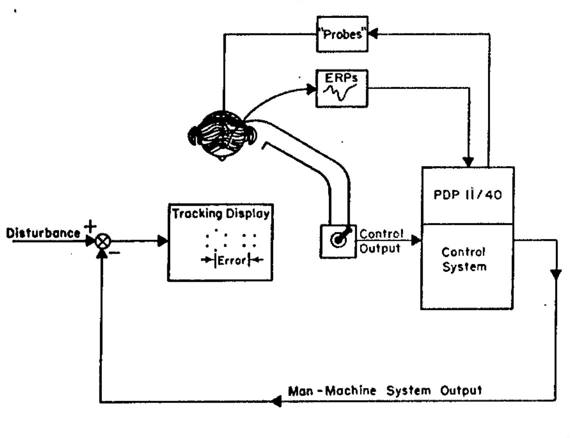


Figure 5. An illustration of the experimental paradigm employed in the analysis of the utility of the ERP as a workload measure.

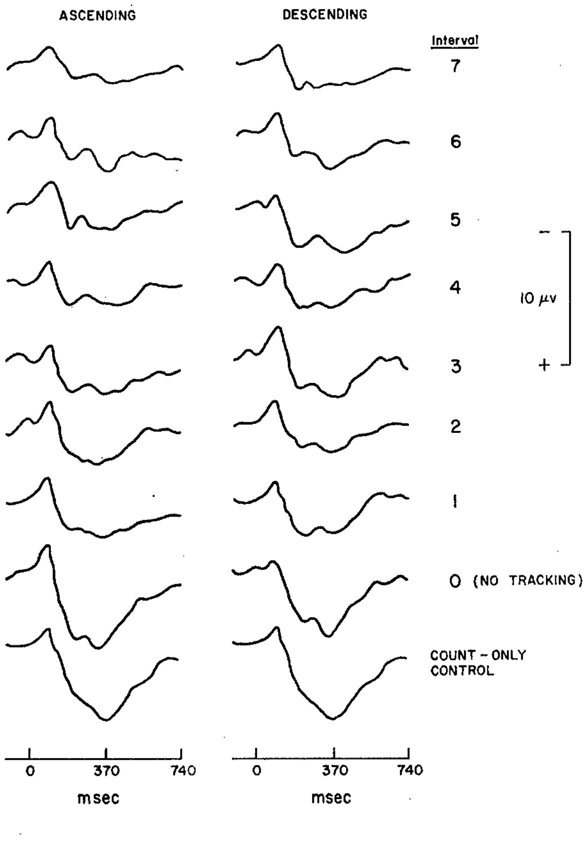


Figure 6. Average parietal ERPs, elicited by equiprobable counted tones, for each bandwidth interval and count-only control conditions, for both ascending and descending blocks of trials (From Isreal, et al., 1980).

These results cannot be explained easily within the framework of an undifferentiated capacity theory if we assume that P300 amplitude indexes the demands placed on the subject by the primary task. Increasing the bandwidth clearly affects the performance of overt secondary tasks (McDonald, 1973; Wierwille, Gutmann, Hicks and Muto, 1977). The fact that P300 did not change, even though a dramatic drop in amplitude was observed with the introduction of the task, required explanation.

One interpretation of the results is that the P300 is not sensitive to the processing demands of the task but instead reflects the motor activity required by tracking. The hypothesis was tested by Isreal et al. (1980) who instructed subjects to manipulate a joystick with one hand concurrently with the oddball task. The amplitude of the P300 component elicited by the tones was not affected by the motor demand. Thus, it would seem that hand movements, per se, did not decrease the amplitude of the P300.

Another interpretation of the results is that the resources that are tapped when the dimensionality, or the bandwidth of the target, are increased are not the resources required by the oddball task. Several investigators have proposed that processing resources are not undifferentiated but are rather structured according to various information processing stages (Kantowitz and Knight, 1976; Kinsbourne and Hicks, 1978; Navon and Gopher, 1979, 1980; Sanders, 1979). Wickens (1980) has identified hypothetical processing structures on the basis of input and output modalities (visual-auditory, manual-vocal), stages of information processing (encoding and central processing, response selection and execution) and codes of processing (verbal, spatial). In this framework dual-tasks are expected to interfere to the extent that they share overlapping resources. For example, two tasks which both require substantial central processing will interfere with each other to a greater extent than a task with central processing demands and another with heavy demands for response processes. This view of the allocation of processing resources is consistent with studies which show little or no decrement in performance when two difficult tasks are time-shared (Allport, Antonis and Reynolds, 1972; North, 1977; Wickens and Kessel, 1979).

The notion that P300 is sensitive to a specific aspect of information processing is consistent with the data, reviewed above, regarding the relation between P300 latency and reaction time. P300 latency appears to be sensitive to a subset of the processes that determine reaction time. Furthermore, P300 latency is influenced by manipulations of factors which are assumed to affect relatively early, stimulus evaluation processes while being insensitve to changes in variables which produce their effect on the later response selection and execution processes. If the manipulation of the dimensionality and bandwidth of the tracking task demand resources associated largely with response selection and execution processes then P300 amplitude should not reflect fluctuations in performance. On the other hand, if the perceptual aspects of a task were manipulated, the amplitude of the P300 elicited by a secondary task would be expected to covary with primary task difficulty.

Isreal, Wickens, Chesney and Donchin (1980) tested the latter hypothesis by combining the oddball task as a secondary task with a visual monitoring task that served as the primary task. The subjects were instructed to monitor a simulated air traffic control display either for course changes or for intensifications of one of two classes of stimuli (triangles or squares). Primary task difficulty was manipulated by increasing the number of elements traversing the CRT (Sperando, 1978). The numerosity variable did in fact have a systematic effect on reaction time to the tones when subjects were monitoring for course changes. Reaction time increased monotonically from the control condition to the condition in which subjects were required to monitor eight elements simultaneously. However, in the flash detection condition reaction time did not increase significantly as a function of the number of elements displayed.

As can be seen from Figure 7 the P300 elicited by the counted tones decreased monotonically with increases in difficulty in the monitoring task when subjects were detecting course changes. In the flash detection condition P300s decreased with the introduction of the monitoring task, but increases in the number of display elements failed to further attenuate P300 amplitude. This result is also consistent with the reaction time data. Since the primary task did not require a response, the data of Isreal et al. (1980) have demonstrated that P300 amplitude is sensitive to the perceptual demands of a primary task.

The Use of P300 in Task Analysis

This structure-specific conception of processing resources has several implications for the study of man-machine systems. One area which might benefit from the use of the structure-specific analysis of human information processing resources is task analysis. Traditionally, the analysis of operator performance in complex systems has been conducted by detailing the observable aspects of tasks and task sequences (Kidd and Van Cott, 1972). This analysis has usually taken the form of elaborate flow charts which outline aspects of operator behavior such as information input, decisions and required actions (Coakley and Fucigna, 1955; Folley, et al., 1960). Although these procedures provide an accurate description of the behavior exhibited by the operators, they do not enable micro analysis of the task that could provide the system designer with information on the resources required by different sub-task sequences. It would be useful to examine a breakdown in performance under high workload conditions for their relation to resource-competition. For

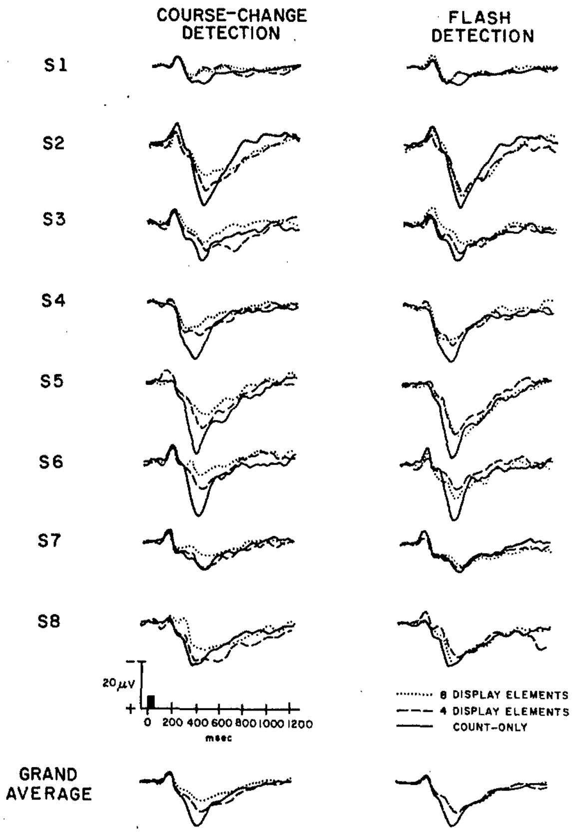


Figure 7. Single subject and average ERPs elicited by infrequent, counted tones presented concurrently with each of two monitoring tasks. Two monitoring conditions as well as a count-only control condition are presented. All waveforms displayed were recorded at Pz (From Isreal, et al., 1980).

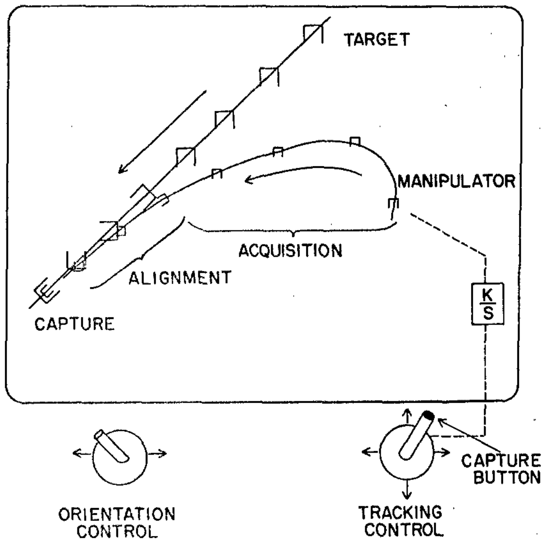


Figure 8. The temporal sequence of the target acquisition task (from upper right to lower left). The large element is the computer controlled target while the smaller element represents the subject controlled cursor. The joy stick on the right side regulates the path of the cursor in the X and Y axes. The joy stick on the left controls the rotational velocity of the cursor (From Wickens, et al., 1982).

example, it would be useful to know if the operator is required to perform tasks which demand a great deal of response processing but little perceptual analysis.

Wickens, Kramer and Donchin (1982) performed a componential analysis of the demands of controlling higher order systems, well validated in the literature to impose a greater load on information processing resources (Baty, 1971; Fuchs, 1962). By "order of control" we refer to the number of time integrations of the output of a controller (i.e., joystick) and the output of the system. In a first order, or velocity driven system, a deflection of the joystick corresponds to a change in the velocity of the controlled element. A second order, or acceleration driven system, produces a change in the accleration of the controlled element proportional to the movement of the control stick. Assuming that P300 amplitude is sensitive to the perceptual aspects of a task then a reduction in P300 amplitude by higher order control should localize the influence of the order variable at the earlier processing stages.

Figure 8 illustrates the subject's task. The target appeared on the screen and moved in a straight line, but at a randomly selected angle, in the direction of its exit. The subject had to move the cursor into the neighborhood of the target. The time between the appearance of the target and its acquisition by the cursor is called the "acquisition phase". Acquisition was accomplished by manipulating the two-axis joystick mounted on the right side of the subject's chair. Successful acquisition initiated the alignment phase. The target began to rotate at a constant velocity in either a clockwise or counterclockwise direction. The subjects had to rotate the cursor at the same velocity as the target while also keeping the two elements superimposed. The rotation was accomplished by manipulating the single axis joystick mounted on the left side of the subject's chair. A deflection of the stick to the right produced a clockwise rotation of the cursor at an angular velocity proportional to the angle of deflection, a deflection to the left produced a counterclockwise rotation. Deviation from the initial acquisition criterion for more than 1000 msec necessitated a re-alignment of the elements. Once the subject decided that all of the criteria had been satisfied and the target and cursor were aligned, she could press a capture button and the trial was terminated.

We assumed that the alignment phase would be more difficult than the acquisition phase due to increased perceptual demands imposed by the requirement to control the additional rotational axis. We predicted, therefore, that the P300 amplitude elicited by the tones, associated with an oddball task run concurrently with the tracking task, would be larger during the acquisition than during the alignment phase.

The ERP results presented in Figure 9 confirm these predictions. The P300 amplitude is attenuated both as a function of phase, larger amplitude P300s being elicited in the acquisition phase, and system order, larger P300s elicited during the easier, first order tracking. Other investigators employing a compensatory tracking task have also found a systematic relationship between P300 amplitude and system order (Wickens, Gill, Kramer, Ross and Donchin, 1981). These studies, along with additive factors investigators of manual control parameters, have provided converging evidence that system order has a salient perceptual/central processing component (Wickens and Derrick, 1981; Wickens, Derrick, Micallizi and Berringer, 1980) The results will also be useful in the design and evaluation of complex tracking tasks. If operators are required to perform a tracking task with higher order system dynamics then concurrently performed tasks should be designed so as to minimize perceptual/central processing load. We see here, again, how the ERPs provide data that increase the theoretical depth with which one can draw conclusions about the human information processing system.

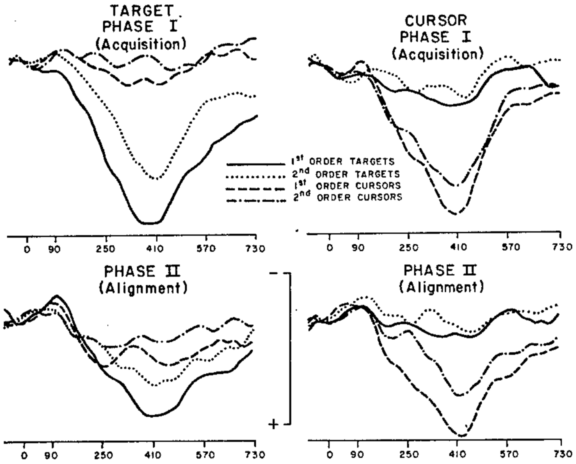


Figure 9. Average parietal waveforms elicited by intensifications of the tracking elements. The left panel presents waveforms recorded when the intensity of the target was the relevant event. The right panel displays waveforms collected when the intensity of the cursor was relevant. The top panels display waveforms recorded during the acquisition phase, the button panels present waveforms collected during the alignment phase of the target acquisition task (From Wickens,' et al., 1982).

P300 and Resource Reciprocity

The studies cited above have demonstrated a robust relationship between P300 amplitude and the allocation of processing resources in a secondary task. P300s elicited by secondary task probes decrease in amplitude with increases in the perceptual/central processing difficulty of primary tasks. As outlined previously, one of the basic assumptions of the secondary task technique is that increases in primary task difficulty divert processing resources from the secondary task. The decrement in secondary task performance is believed to reflect this shift of resources from the secondary to the primary task. Thus, it is assumed that there is a reciprocal relationship between the resources allocated to the primary and secondary tasks. If this assumption is correct, then it should be possible to demonstrate that P300s elicited by task relevant, discrete events embedded within the primary task are directly related to primary task difficulty.

Kramer, Wickens, Vanasse, Heffley and Donchin (1981) conducted an experiment in which ERPs were elicited by task relevant events embedded within a pursuit step tracking task. The subjects were required to perform a single axis pursuit step tracking task with either first order (velocity) or second order (acceleration) control dynamics. In this task the horizontal position of a target was determined by a random series of step displacements occurring at 3 sec intervals. The subjects task was to keep the cursor superimposed on the target. Difficulty was varied by manipulating two variables: the degree of predictability of the series of steps and the system order. In the high predictability condition the step changes alternated in a regular right-left pattern. In the low predictability condition the sequence of step changes was random. The magnitude of the changes was unpredictable in both conditions. The two dimensions of difficulty, system order and input predictability, were crossed to create three conditions of increasing difficulty: first order control of predictable input, first order control of unpredictable input and second order control of unpredictable input.

Three different types of probes were employed as ERP eliciting events. In one condition subjects performed the tracking task while also counting the number of occurrences of a low pitched tone from a Bernoulli series of high and low pitched tones. In the second condition subjects counted the dimmer of two flashes in a Bernoulli sequence. The flash appeared as a horizontal bar along the path traversed by the target. In the primary task probe condition, subjects counted the total number of step changes to the left. Two control conditions were also included: one in which the subjects counted the probes but did not track and a second in which subjects performed the tracking task without counting the probes.

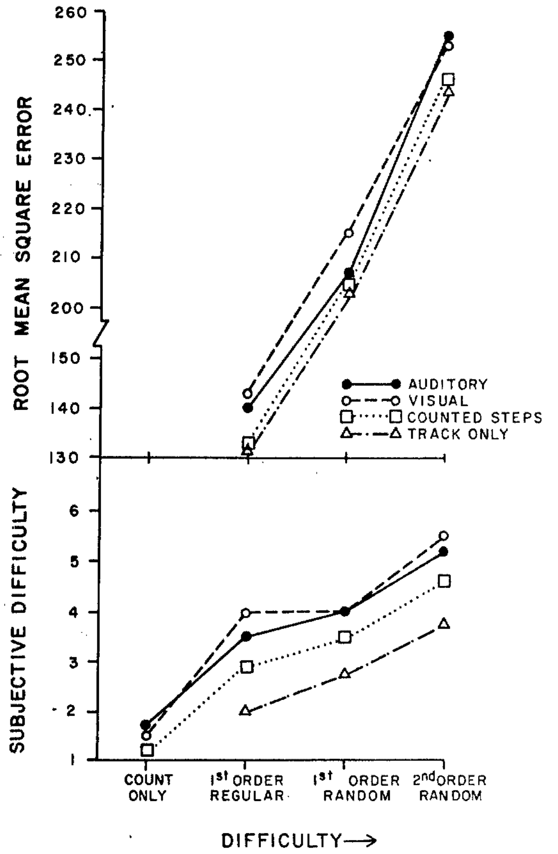


Figure 10. Average Root Mean Square error and subjective difficulty ratings recorded for each condition in a pursuit step tracking task (After Kramer, et al., 1981).

The important findings to note in the data presented in Figure 10 are the monotonic relations between the tracking difficulty manipulations and the subject's perceived ratings of difficulty, as well as those between tracking difficulty and RMS error. Thus both the subjective and behavioral indices converge on the same ordering of task difficulty. However, these measures do not provide information concerning the underlying resource structure of the task.

The effect of tracking difficulty on P300 amplitude in the auditory condition provide results consistent with previous research (Isreal, et al., 1980; Wickens, et al., 1980). Thus, in the auditory condition, an increase in the difficulty of the primary task decreased the amplitude of the P300 elicited by the secondary task probes. In the visual condition the introduction of the tracking task resulted in a reduction in the amplitude of the P300. However, increases in tracking difficulty failed to produce any further attenuation. In the step conditions, the amplitude of the P300 elicited by the discrete changes in the spatial position of the controlled element increased with increments in the difficulty of the primary task. Thus, the hypothesis of resource reciprocity between the primary and secondary tasks was confirmed. One final aspect of the step tracking study has considerable potential practical utility. The sensitivity of the P300 elicited by visual steps to resource allocation was observed independent of whether or not the subjects were required to count the stimuli. These data suggest that inferences from the P300 about resource allocation and therefore workload can be made in the total absence of a secondary task requirement, a considerable advantage if workload is to be assessed unobtrusively in real-time environments.

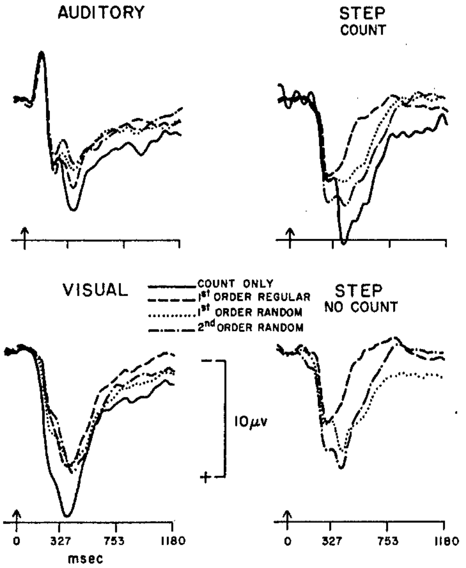


Figure 11. Average parietal ERPs elicited by visual, auditory and spatial probes presented concurrently with a pursuit step tracking task at each level of difficulty. Also shown are the ERPs elicited during count-only control conditions (After Kramer, et al., 1981).

SUMMARY AND CONCLUSIONS

The investigations reported above demonstrate conclusively that the P300 as a secondary task can diagnostically reflect primary task workload variations of a perceptual/cognitive nature, uncontaminated by response factors. The absence of overt response requirements provide it with a considerable advantage over the secondary task, in that the oddball count task is considerably less intrusive.

As a secondary task however, the probe task is not entirely unobtrusive and interpretation of the measures still requires the investigator to make certain assumptions about the nature of the primary-secondary task interaction in order to make inferences concerning operator workload. It is for this reason that our most recent observations that P300 elicited by primary task stimuli also reflect resource allocation are particularly encouraging to the utility of the ERP as a measure of workload in extra-laboratory environments.

We reviewed in this chapter, studies of the ERP that have, we think, one characteristic in common. In each case the ERP served as a source of information on the timing or the "intensity" of an information processing activity whose behavior is not easily monitored by means of observations on overt responses. It would seem that a science of Human Engineering that is interested in developing and testing hypotheses about the internal structure and the operating modes of the human operator would benefit from this additional information. We advocate here the use of the ERP as an analytical tool that can usefully aid in deepening our understanding, and the measurement of, mental workload.

ACKNOWLEDGEMENTS

The work conducted at the Cognitive Psychophysiology Laboratory would not have been possible without the assistance of numerous members of the scientific and technical staff. We would like to thank Mike Anderson, Brian Foote, Dave McWilliams and Ron Clapman for their invaluable programming and engineering help; Diana Mitchell for providing administrative support, Barbara Mullins who processed our words into a manuscript, and Debra O'Huallachain for her graphics assistance. Members of the scientific staff which have been instrumental in the research reported here include Jack Isreal, Daniel Gopher, Greg Chesney, Connie Duncan-Johnson, Linda Vanasse, John Polich, Greg McCarthy and Earle Heffley.

The research has been supported under the Office of Naval Research (Contract #N00014-76-C-0002) with funds provided by the Defense Advanced Research Projects Agency; AFOSR under Contract #F49620-79C-0233; Wright-Patterson Air Force Base under Contract #F33615-79C-0512; and the Environmental Protection Agency under Contract #R805628010. The support of Drs. Craig Fields, Judith Daly, Al Fregley, Don Woodward, Bob O'Donnell and Dave Otto is gratefully acknowledged.

Address requests for reprints to: Emanuel Donchin, Cognitive Psychophysiology Laboratory, Department of Psychology, University of Illinois, Champaign, Illinois, 61820.

REFERENCES

Allport, D.A., Antonis, B., and Reynolds, P. On the division of attention: A disproof of the single channel hypothesis. Quarterly Journal of Experimental Psychology, 1972, 24, 225-235.

Baty, D. Human transformation rates in one to four axis tracking. Proceedings of the 7th Annual NASA Conference on Manual Control, NASA, SP A281, 1971.

Bertelson, P. S-R relationships and reaction time to new versus repeated signals in a serial task. Journal of Experimental Psychology, 1963, 65, 478-484.

Bostock, H. and Jarvis, M.J. Changes in the form of the cerebral evoked response related to the speed of simple reaction time. Electroencephalography and Clinical Neurophysiology, 1970, 29, 137-145.

Brown, I.D. Dual task methods of assessing workload. Ergonomics, 1978, 21, 221-224.

Coakley, J.D. and Fucigna, J.T. Human engineering recommendations for the instrumentation of radar. Report No. AN/FPS-16 (Xn - 2), Radio Corporation of America, Moorestown, New Jersey, 1955.

Clynes, M. and Kohn, M. The use of Mnemotron for biological data storage, reproduction, and for an average transient computer. Abstracts of the 4th Annual Meeting of the Biophysics Society, Philadelphia, Pennsylvania, 1960.

Conrad, C. Some factors involved in the recognition of words. In J.W. Cotton and R.W. Klatzky (Eds.), Semantic Factors in Cognition, Hillsdale, New Jersey: Lawrence Erlbaum, 1978.

Donchin, E. Brain electrical correlates of pattern recognition. In G.F. Inbar (Ed.), Signal Analysis and Pattern Recognition in Biomedical Engineering. New York: John Wiley, 1975.

Donchin, E. Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), Evoked Brain Potentials and Behavior. New York: Plenum Press, 1979.

Donchin, E. Surprise! ... Surprise? Psychophysiology, 1981, 18, 493-513.

Donchin, E. (Ed.) Advances in Cognitive Psychophysiology. Proceedings of the Carmel I Workshop. New York: Erlbaum, in press.

Donchin, E., Kubovy, M., Kutas, M., Johnson, Jr., R. and Herning, R.I. Graded changes in evoked response (P300) amplitude as a function of cognitive activity. Perception and Psychophysics, 1973, 14, 319-324.

Donchin, E., Ritter, W., and McCallum, C. Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, and S. Koslow (Eds.), Brain Event-Related Potentials in Man. New York: Academic Press, 1978.

Donchin, E. and Isreal, J.B. Event-related potentials and psychological theory. In H.H. Kornhuber and L. Deecke (Eds.), Motivation Motor and Sensory Processes of the Brain: Electrical Potentials, Behavioral and Clinical Use. Netherlands: North Holland Biomedical Press, 1980.

Donchin, E., McCarthy, G., Kutas, M., Ritter, W. and Donchin, E. Event-related brain potentials in the study of consciousness. In Davidson, Schwartz and Shapiro (Eds.), Consciousness and Self Regulation. New York: Plenum Press, 1981.

Duncan-Johnson, C.C. and Donchin, E. On quantifying surprise: The variation in event-related potentials with subjective probability. Psychophysiology, 1977, 14, 456-467.

Folley, Jr., J.D., Altman, J.W., Graser, R., Preston, H.O. and Weislogel, R.C. Human factors methods for system design. Report No. AIR-B90-60-FR-225. The American Institutes for Research, Pittsburgh, Pennsylvania, 1960.

Ford, J.M., Mohs, R.C., Pfefferbaum, A. and Kopell, B.S. On the utility of P300 latency and reaction time for studying cognitive processes. In H.H. Kornhuber and L. Deecke (Eds.), Motivation, Motor and Sensory Processes of the Brain: Electrical Potentials, Behavioral and Clinical Use. Netherlands: North Holland Biomedical Press, 1980.

Ford, J.M., Roth, W.T., Mohs, R.C., Hopkins, W.F. and Kopell, B.S. Event-related potentials recorded from young and old adults during a memory retrieval task. Electroencephalography and Clinical Neurophysiology, 1979, 47, 450-454.

Fuchs, A. The progression-regression hypothesis in perceptual-motor skill learning. Journal of Experimental Psychology, 1962, 63, 177-182.

Goff, W.R., Allison, T. and Vaughan, Jr., H.G. The functional neuroanatomy of the event-related potentials. In E. Callaway, P. Tueting, and S. Koslow (Eds.), Brain Event-Related Potentials in Man. New York: Academic Press, 1978.

Gomer, F.E., Spicuzza, R.J., and O'Donnell, R.D. Evoked potential correlates of visual item recognition during memory scanning tasks. Physiological Psychology, 1976, 4, 61-65.

Heffley, E., Wickens, C.D. and Donchin, E. Intramodality selective attention and P300 - reexamination in a visual monitoring task. Psychophysiology, 1978, 15, 269-270.

Hirst, W., Spelke, E.S., Reaves, C.C., Caharack, G. and Neisser, U. Dividing attention without alternation or automaticity. Journal of Experimental Psychology: General, 1980, 109, 98-117.

Isreal, J.B. Structural interference in dual-task performance: Event-related potential, behavioral and subjective effects. Unpublished Doctoral Dissertation, University of Illinois, 1980.

Isreal, J.B., Chesney, G.L., Wickens, C.D. and Donchin, E. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. Psychophysiology, 1980, 17, 259-273.

Isreal, J.B., Wickens, C.D., Chesney, G.L., and Donchin, E. The event-related brain potential as an index of display monitoring workload. Human Factors, 1980, 22, 212-224.

Isreal, J.B., Wickens, C.D., and Donchin, E. The dynamics of P300 during dual-task performance. In H.H. Kornhuber and L. Deecke (Eds.), Motivation, Motor and Sensory Processes in the Brain: Electrical Potentials, Behavioral and Clinical Use. Netherlands: North Holland Biomedical Press, 1980.

Jewett, D., Romano, H.W., and Williston, J.S. Human auditory evoked responses: Possible brain stem components detected on the scalp. Science, 1970, 167, 1517-1518.

Johanssen, G., Moray, N., Pew, R., Rasmussen, J., Sanders, A., and Wickens, C. Final report of the experimental psychology group. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement. New York: Plenum Press, 1979.

Johnson, Jr., R.E. and Donchin, E. On how P300 amplitude varies with the utility of the eliciting stimuli. Electroencephalography and Clinical Neurophysiology, 1978, 44, 424-437.

Kahneman, D. Attention and Effort. Englewood Cliffs, N. J.: Prentice Hall, 1973.

Kantowitz, B.H. and Knight, J.L. On experimenter limited processes. Psychological Review, 1976, 83, 502-507.

Karlin, L., Martz, M.J., and Mordkoff, A.M. Motor performance and sensory evoked potentials. Electroencephalography and Clinical Neurophysiology, 1970, 28, 307-313.

Karlin, L. and Martz, M.J. Response probability and sensory evoked potentials. In S. Kornblum (Ed.), Attention and Performance IV. New York: Academic Press, 1973.

Kidd, J.S. and Van Cott, H.P. System and human engineering analysis. In H.P. Van Cott and R.G. Kinkade (Eds.), Human Engineering Guide to Equipment Design, Washington, D.C.: U.S. Government Printing Office, 1972.

Kinsbourne, M. and Hicks, R.E. Functional cerebral space: A model for overflow, transfer and interference effects in human performance. In J. Requin (Ed.), Attention and Performance VII. New York: Academic Press, 1978.

Knowles, W.B. Operator loading tasks. Human Factors, 1963, 5, 155-161.

Kramer, A. and Donchin, E. A chronometric analysis of the role of orthographic and phonological cues in a non-lexical decision task, Psychophysiology, (Abstract), in press, 1982.

Kramer, A., Wickens, C.D., Vanasse, L., Heffley, E.F. and Donchin, E. Primary and secondary task analysis of step tracking: An event-related potentials approach. In R.C. Sugarman (Ed.), Proceedings of the 25th Annual Meeting of the Human Factors Society, Rochester, New York, 1981.

Kutas, M., McCarthy, G., and Donchin, E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 1977, 197, 792-795.

McCarthy, G. and Donchin, E. A metric for thought: A comparison of P300 latency and reaction time. Science, 1981, 211, 77-80.

McClelland, J.L. On time relations of mental processes: A framework for analyzing processes in cascade. Psychological Review, 1979, 86, 287-330.

McDonald, L.B. A model for predicting driver workload in the freeway environment: A feasibility study. Unpublished Doctral Dissertation, Texas A and M University, 1973.

Meyer, D.E., Schvaneveldt, R.W. and Ruddy, M.G. Loci of contextual effects on visual word recognition. In P. Rabbit and S. Dornic (Eds.), Attention and Performance V. New York: Academic Press, 1975.

Moray, N. Where is capacity limited? A survey and a model. In A.F. Sanders (Ed.), Attention and Performance I. Amsterdam: North Holland, 1967.

Natani, K. and Gomer, F.E. Electrocortical activity and operator workload: A comparison of changes in the Electroencephalogram and in event-related potentials. McDonnell Douglas Technical Report, MDC E2427, McDonnell Douglas Astronautics Company, St. Louis, Missouri, June, 1981.

Navon, D. and Gopher, D. On the economy of the human processing system. Psychological Review, 1979, 86, 214-255.

Navon, D. and Gopher, D. Interpretations of task difficulty in terms of resources: Efficiency, load, demand and cost composition. In R. Nickerson and R. Pew (Eds.), Attention and Performance VIII. Hillsdale, New Jersey: Erlbaum, 1980.

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

North, R. Task functional demands as factors in dual-task performance. Proceedings of the 21st Annual Meeting of the Human Factors Society, Santa Monica, California, 1977.

Ogden, G.D., Levine, J.W. and Eisner, E.J. Measurement of workload by secondary tasks. Human Factors, 1979, 21, 529-548.

Otto, D.A. (Ed.) Multidisciplinary Perspectives in Event-Related Brain Potential Research, EPA 600/9-77-043, Washington, D.C.: U.S. Government Printing Office, 1979.

Picton, T.W. and Stuss, P.T. The component structure of the human event-related potentials. In H.H. Kornhuber and L. Deecke (Eds.), Motivation, Motor and Sensory Processes of the Brain: Electrical Potentials, Behavioral and Clinical Use. Netherlands: North Holland Biomedical Press, 1980.

Posner, M.I. Chronometric Explorations of the Mind. Hillsdale, N. J.: Erlbaum, 1978.

Rasmussen, J. Models of mental strategies in process plant diagnosis. In J. Rasmussen and W.B. Rouse (Eds.), Human Detection and Diagnosis of System Failures. New York: Plenum Press, 1981.

Regan, D. Evoked Potentials in Psychology, Sensory Physiology and Clinical Medicine. London: Chapman and Hall, 1972.

Ritter, W., Simson, R., Vaughan, Jr., H.G. Association cortex potentials and reaction time in auditory discriminations. Electroencephalography and Clinical Neurophysiology, 1972, 33, 547-557.

Rohrbaugh, J.W., Donchin, E., and Ericksen, C.W. Decision making and the P300 component of the cortical evoked response. Perception and Psychophysics, 1974, 15, 368-374.

Rolfe, J.M. The secondary task as a measure of mental load. In W.T. Singleton, J.G. Fox and D. Witfield (Eds.), Measurements of Man at Work. London: Taylor and Francis, 1971.

Ruchkin, D.S. and Sutton, S. Visual evoked and emitted potentials and stimulus significance. Psychonomic Society Bulletin, 1973, 2, 144-146.

Sanders, A.F. Some variables affecting the relation between relative stimulus frequency and choice reaction time. In A.F. Sanders (Ed.), Attention and Performance III. Amsterdam: North Holland, 1970.

Sanders, A.F. Some remarks on mental load. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement. New York: Plenum Press, 1979.

Satterfield, J.H. Evoked cortical response enhancement and attention in man: A study of responses to auditory and shock stimuli. Electroencephalography and Clinical Neurophysiology, 1965, 19, 470-475.

Schneider, W. and Shiffrin, R.M. Controlled and automatic human information processing: I. Detection, search and attention. Psychological Review, 1977, 84, 1-66.

Sheridan, T.B. and Simpson, R.W. Toward the definition and measurement of mental workload of transport pilots. Cambridge, Massachusetts: Massachusetts Institute of Technology Flight Transportation Laboratory Report, FTL Report R79-4, January, 1979.

Sheridan, T.B. Mental Workload - What is it? Why bother with it? Human Factors Society Bulletin, 1980, 23, 1-2.

Sheridan, T.B. Understanding human error and aiding human diagnostic behavior in nuclear power plants. In J. Rasmussen and W.B. Rouse (Eds.), Human Detection and Diagnosis of System Failures. New York: Plenum Press, 1981.

Shiffrin, R.M. and Schneider, W. Controlled and automatic human information processing II. Perceptual learning, automatic attending and a general theory. Psychological Review, 1977, 84, 127-190.

Shwartz, S.P., Pomerantz, J.R., and Egeth, H.E. State and process limitations in information processing: An additive factors analysis. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 402-410.

Sokolov, E.W. The modeling properties of the nervous system. In I. Maltzman and K. Cole (Eds.), Handbook of Contemporary Soviet Psychology. New York: Basic Books, 1969.

Spelke, E.S., Hirst, W.C. and Neisser, W. Skills of divided attention. Cognition, 1976, 4, 215-230.

Sperando, J.C. The regulation of working methods as a function of workload among air traffic controllers. Ergonomics, 1978, 21, 193-202.

Squires, N.K., Donchin, E., Squires, K.C., and Grossberg, S. Bisensory stimulation: Infering decision related processes from the P300 component. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 299-315.

Starr, A., Sohmer, H. and Celesia, G.G. Some applications of evoked potentials with neurological and sensory impairment. In E. Callaway, P. Tueting, and S. Koslow (Eds.), Brain Event-Related Potentials in Man. New York: Academic Press, 1978.

Sternberg, S. High speed scanning in human memory. Science, 1966, 153, 652-654.

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. Acta Psychologica, 1969a, 30, 276-315.

Sternberg, S. Memory Scanning: Mental processes revealed by reaction time experiments. American Scientist, 1969b, 57, 421-457.

Stuss, D.T., and Picton, T.W. Neurophysiological correlates of human concept formation. Behavioral Biology, 1978, 23, 135-162.

Sutton, S., Braren, M., Zubin, J. and John, E.R. Evoked potential correlates of stimulus uncertainty. Science, 1965, 150, 1187-1188.

Sutton, S., Tueting, P., Zubin, J., and John, E.R. Information delivery and the sensory evoked potential. Science, 1967, 155, 1436-1439.

Wickelgren, W. Speed accuracy tradeoff and information processing dynamics. Acta Psychologica, 1977, 41, 67-85.

Wickens, C.D. Human workload measurement. In N. Moray (Ed.), Mental Workload: Its Theory and Measurement. New York: Plenum Press, 1979.

Wickens, C.D. The structure of attentional resources. In R. Nickerson and R. Pew (Eds.), Attention and Performance VIII, Hillside, N. J.: Erlbaum, 1980.

Wickens, C.D., Isreal, J.B., and Donchin, E. The event-related cortical potential as an index of task workload. Proceedings of the 21st Annual Meeting of the Human Factors Society, Santa Monica, California, 1977.

Wickens, C.D. and Derrick, W. The processing demands of higher order manual control: Application of additive factors methodology. Engineering-Psychology Research Laboratory, University of Illinois, EPL-81-1/ONR-81-1, 1981.

Wickens, C.D., Derrick, W.D., Micallizi, J., and Berringer, D. The structure of processing resources. In R.E. Corrick, E.C. Haseltine and R.T. Durst (Eds.), Proceedings of the 24th Annual Meeting of the Human Factors Society, Los Angeles, California, 1980.

Wickens, C.D., Gill, R., Kramer, A., Ross, W. and Donchin, E. The cognitive demands of second order manual control: Applications of the event-related brain potential. Proceedings of the 17th Annual NASA Conference on Manual Control, NASA TM, 1981.

Wickens, C.D., Heffley, E.F., Kramer, A.F., and Donchin, E. The event-related brain potential as an index of attention allocation in complex displays. In G.E. Corrick, E.C. Haseltine and R.T. Durst (Eds.), Proceedings of the 24th Annual Meeting of the Human Factors Society, Los Angeles, California, 1980.

Wickens, C.D., and Kessel, C. The effect of participatory mode and task workload on the detection of dynamic system failures. IEEE Transactions on Systems, Man and Cybernetics, 1979, SMC-13, 24-34.

Wickens, C.D., Kramer, A.F., and Donchin, E. The event-related potential as an index of the processing demands of a complex target acquisition task. In R. Karrer (Ed.), Proceedings of the Sixth International Conference on the Event-Related Slow Potentials of the Brain, in press, 1982.

Wickens, C.D., Kramer, A.F., Vanasse, L. and Donchin, E. The reciprocity of primary and secondary task resources: Evidence from the P300 component of the ERP, submitted, 1982.

Wierwille, W.W., Gutmann, J.C., Hicks, T.G., and Muto, W.H. Secondary task measurements of workload as a function of simulated vehicle dynamics and driving conditions. Human Factors, 1977, 19, 557-566.

Wilkinson, R.T. and Morlock, Jr., H.C. Auditory evoked response and reaction time. Electroencephalography and Clincial Neurophysiology, 1967, 23, 50-56.

Williges, R.C. and Wierwille, W.W. Behavioral measures of aircrew mental workload. Human Factors, 1979, 21, 549-574.

Wood, C.C. and Allison, T. Interpretation of evoked potentials: A neurophysiological perspective. Canadian Journal of Psychology, 1981, 35, 113-135.

Woody, C.D. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Medical and Biological Engineering, 1967, 5, 539-553.

SPEECH-RELATED BRAIN POTENTIALS ELICITED BY SYNTHETIC SPEECH STIMULI

M. A. Biferno, Ph.D.
Douglas Aircraft Company
Long Beach, California

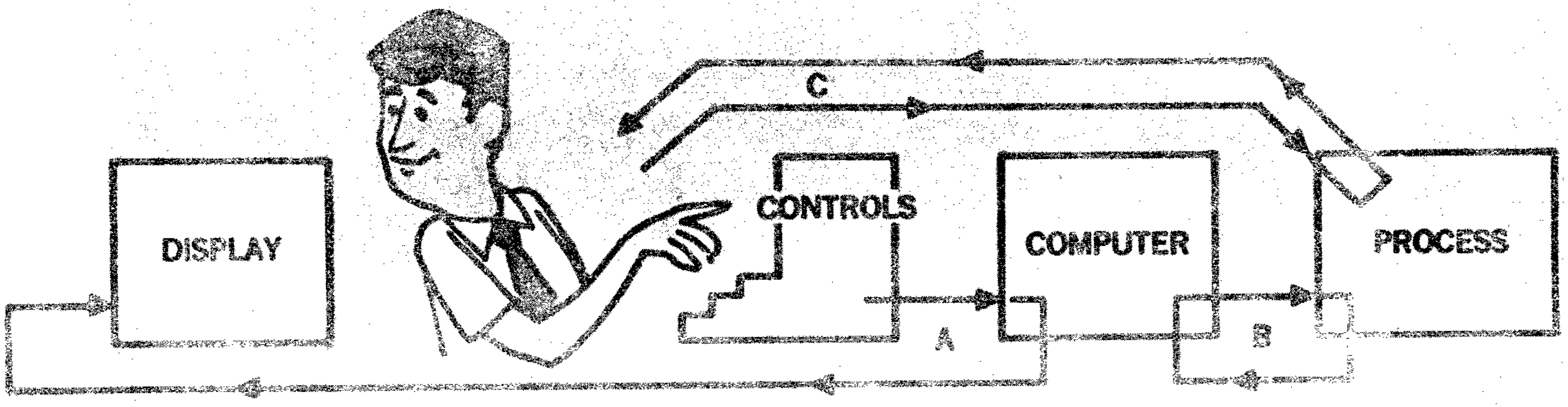
Abstract

Since synthetic speech stimuli are currently employed in commercial aircraft systems, this stimulus type has promise as a relatively unobtrusive probe to elicit event-related brain potentials (ERP) while people perform mental work in an operational environment. Experimental work has shown the feasibility of recording speech-related potentials from synthetic speech stimuli while employing standard techniques. Preliminary results have shown reliable increases in the P300 component at the Pz site when subjects perform mental work to a word stimulus when compared to conditions where the word stimulus is being ignored.

The ERP data has also been processed for display on a color graphics CRT. The changes in scalp voltage distributions show clearly differentiated patterns of activity as a person senses, attends, and mentally processes the synthetic speech stimuli we employed in our experiment. The interpretation of these color patterns is based on current understandings of brain neuroanatomy and ERP theory.

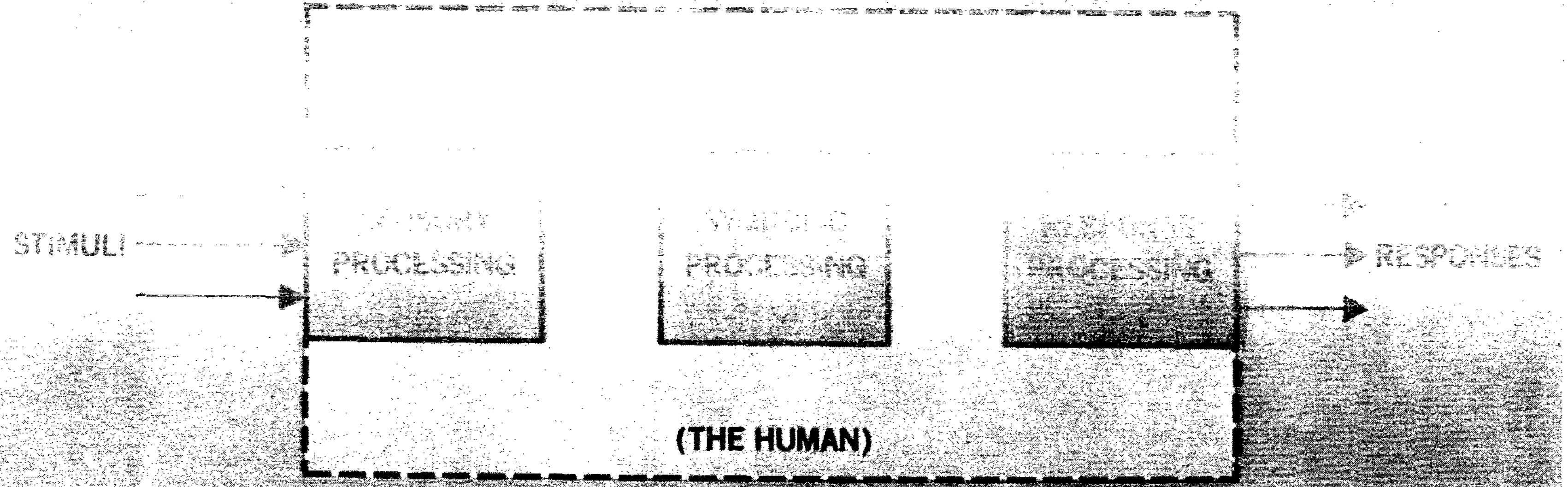
We are studying the brain potential activity associated with mentally processing language stimuli because it is anticipated that much of the mental workload of future automated systems will be languaged-based thinking. (For example; recalling procedures and system commands, formulating verbal communications and strategies, solving problems, and isolating system faults.) If we could measure when the brain was engaged in language-based processing, then we might have a useful index of mental workload which could be employed as a design aid for complex automated systems.

MAN/COMPUTER INTERACTION



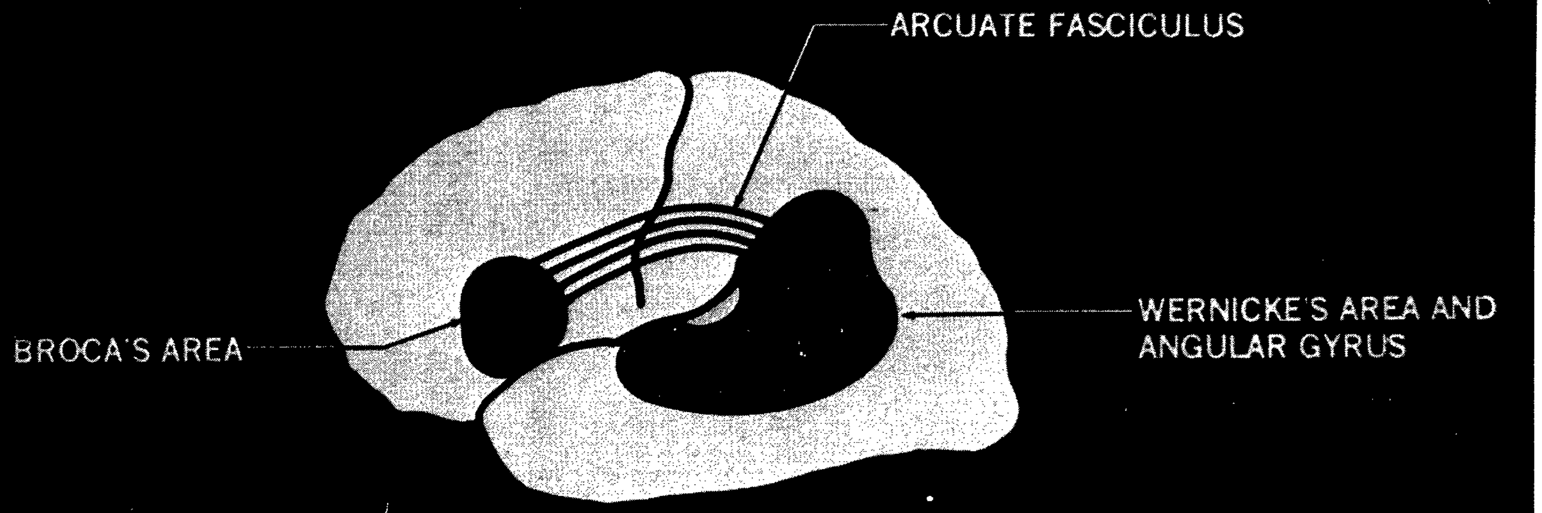
80-GEN-22507

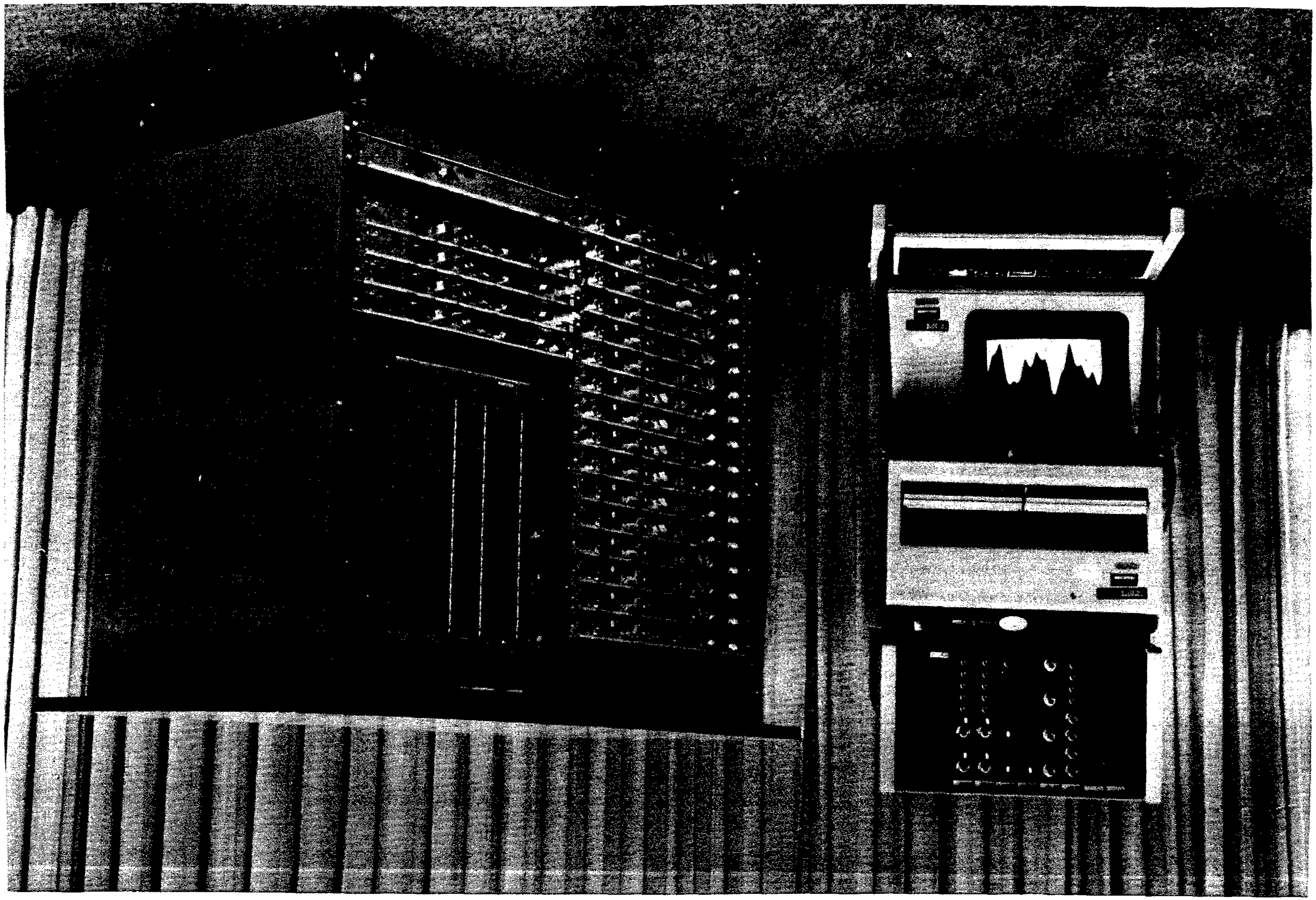
HUMAN INFORMATION PROCESSING APPROACH



9-GEN-28659 A

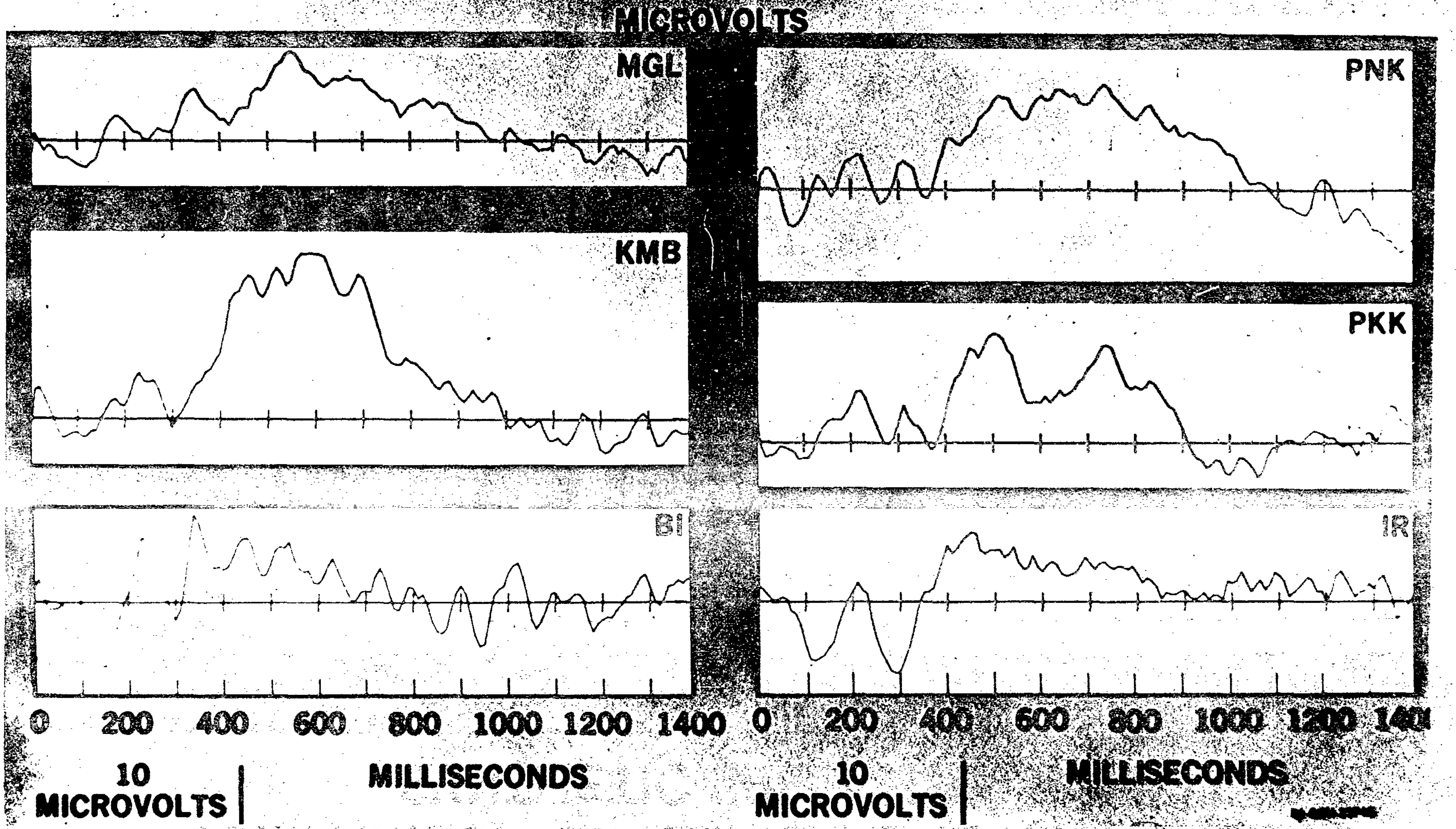
LANGUAGE AREAS



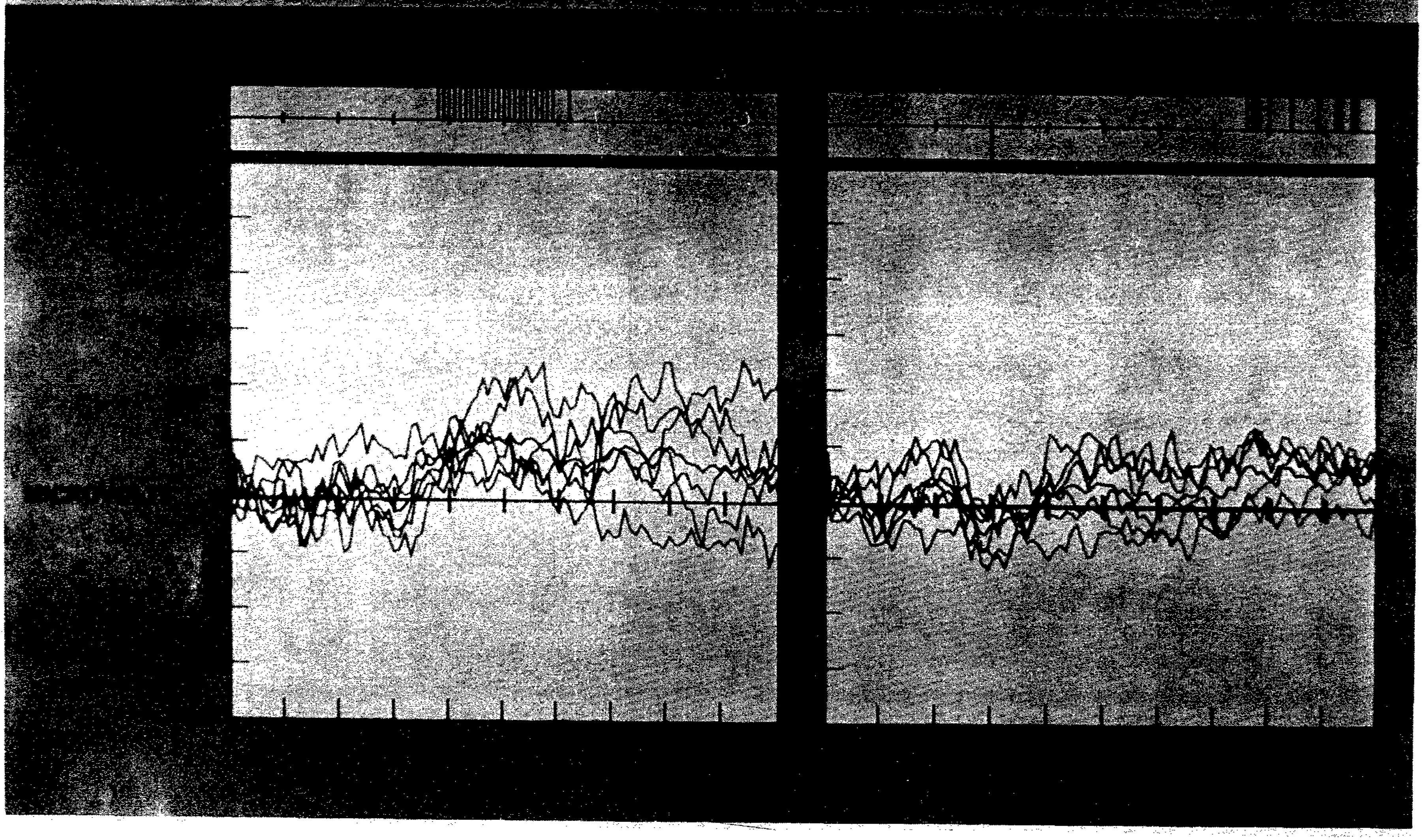


SPEECH-RELATED POTENTIALS

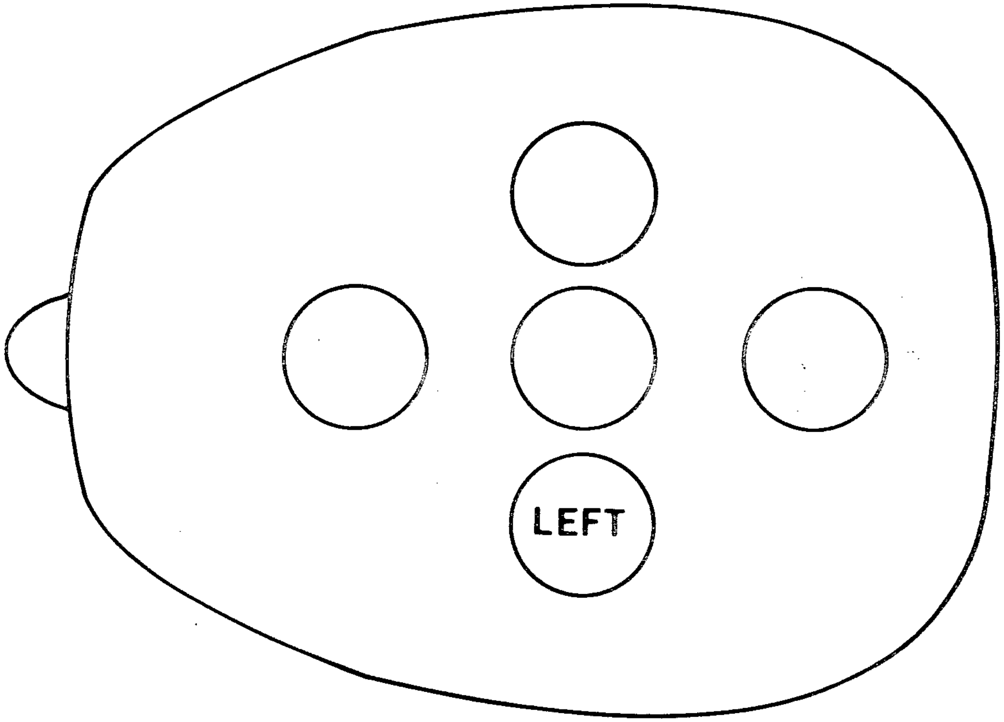
SITE: Pz



COMPARISON OF [illegible]
SYNTHETIC [illegible]



SENSOR LOCATIONS



NEW TOOLS FOR ASSESSING AIRCRAFT/PILOT PERFORMANCE

By: R. L. York
L. D. Montgomery
J. P. Petro

SRI International
Menlo Park, California

Presented at:

AIAA Workshop
Flight Testing to Identify Pilot Workload
and Pilot Dynamics

19-21 January 1982

Air Force Flight Test Center
Edwards Air Force Base, California

I INTRODUCTION

A. Background

Today's high-performance tactical aircraft are sophisticated, state-of-the-art weapon systems consisting of three components: the aircraft, the pilot, and the aircraft's armament. Successful use of these systems requires that each component perform satisfactorily, both individually and in conjunction with each other. As aircraft and armament subsystems evolve, pilot functions also change. For example, as aircraft performance expands (greater speeds, larger g-loads), the pilot is exposed to a more stressful physical environment. Pilot aids (such as a reclined seat, antigravity suit, and side-stick control) were developed to counter this stressful environment. The pilot's mental workload includes not only flying the aircraft, evaluating the tactical situation in combat, and utilizing the correct displays and controls, but also selecting the proper weapon from the armament suite. Many new sophisticated weapons are multimode, causing further stress on the pilot by burdening him with mode selection. The data and information processing workload must be divided between the pilot and onboard computers, and the point at which this division is made significantly affects the stress the pilot must accommodate.

A multidimensional approach is necessary to measure pilot workload, stress, and performance quantitatively. These measurements require the simultaneous acquisition of aircraft performance, pilot physiologic response, and pilot performance data during tactical training flight maneuvers. With these data, the overall performance characteristics of the weapon system can be assessed in a realistic air combat situation. These data may then be used in evaluating weapon system design and pilot training, in predicting pilot response to stressful environments, and in developing countermeasures to enhance pilot capabilities.

Tools are now available to evaluate both aircraft and pilot performance individually, but they have never been combined to test the complete weapon system (which includes the pilot) in a realistic air combat situation. Today's technology is capable of evaluating aircraft/pilot performance in a cost-effective manner. SRI is currently involved in assessing all three areas of aircraft weapon system effectiveness (aircraft performance, pilot physiological response, and aircraft armament utilization) and is able to correlate and analyze these assessments to determine their unique relationships.

B. Objective

This paper presents SRI's current capability for gathering both aircraft and armament performance data and simultaneous pilot physiological response data. It also explores the potential near-future expansion of this capability. Section II describes SRI's aircraft and pilot instrumentation systems. Section III outlines current performance measurement capabilities, and Section IV describes the expanded measurements that will be possible in the near future. Finally, Section V summarizes the advantages of SRI's multidimensional approach.

II SYSTEM DESCRIPTIONS

A. TACTS/ACMI

SRI is technical advisor to the Department of Defense for development of an inflight aircrew training system called Tactical Aircrew Combat Training System (TACTS) by the U.S. Navy and Air Combat Maneuvering Instrumentation (ACMI) by the U.S. Air Force. The system features (1) real-time tracking, integration, processing, and display of maneuvering aircraft and associated flight data, (2) computer simulation of weapon system employment against which aircrews can exercise their ability in flight, and (3) magnetic recording of exercise data for subsequent replay and in-depth analysis.

Today, five TACTS/ACMI systems (Figure 1) are in operation, training aircrews in air-to-air combat. The TACTS/ACMI arena for air-to-air combat training is approximately 60 to 80 km in diameter and 15 to 20 km in altitude. In this arena, opposing aircrews in supersonic aircraft engage in free-play mock combat in an environment almost identical to that of real combat. Range instrumentation tracks the aircraft in real time in position, velocity, acceleration, and attitude; it also provides computerized simulations of air-to-air missiles. Instrumentation carried in a pod on each participating aircraft relays pilot-initiated missile-firing signals to ground-based computers to initiate the missile simulations. Target tracking data recorded by the range system are used in the simulations to determine whether the missile would have resulted in a kill, had a real missile been fired. The results can be immediately communicated to the pilot by the Range Training Officer via ground-to-air radio. Range activity is displayed in real time, in three-dimensional perspective. All data are recorded on magnetic tape, permitting replay during debriefing of all range activities.

In brief, TACTS/ACMI offers three generic features that underlie both its value as a training/test system and its versatility:

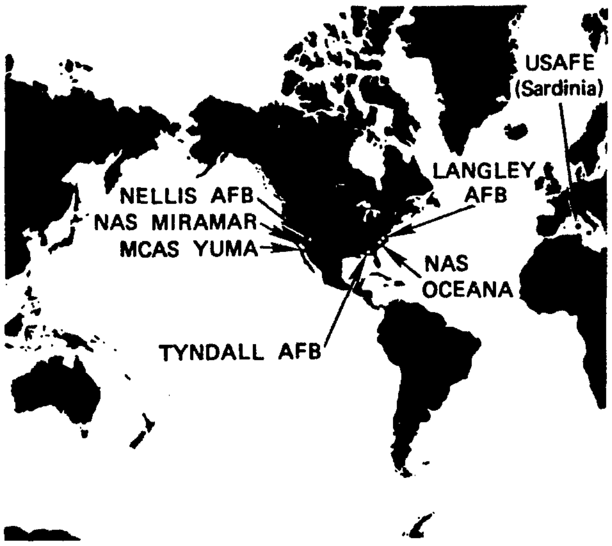


FIGURE 1 TACTS/ACMI INSTALLATION LOCATIONS

- Real-time integration, processing, and display of actual aircraft maneuvers and flight data ("see it now").
- Computer simulations of weapon system employment against which aircrews can exercise their abilities in flight.
- Magnetic recording of exercise data for subsequent replay of alphanumeric and 3-D graphic displays ("see it later").

Basically, the TACTS/ACMI is an integrated tracking and data transmission system consisting of four major subsystems: the Airborne Instrumentation Subsystem (AIS), the Tracking Instrumentation Subsystem (TIS), the Control and Computation Subsystem (CCS), and the Display and Debriefing Subsystem (DDS). The system operates as a closed-loop, state-vector tracking system by combining in the CCS the attitude and angle-rate data from the AIS with multilateration range-tracking data from the TIS.

1. Airborne Instrumentation Subsystem

Each aircraft operating on the TACTS/ACMI range carries an AIS pod (Figure 2). The AIS pod, which is similar in size and shape to a Sidewinder missile, contains a transponder, an inertial sensor assembly, a digital processor, a digital interface unit, and an air data sensor probe and transducer. The AIS pod is designed to be carried on any aircraft station that can carry the Sidewinder (AIM-9) and can be attached in less than 5 minutes. The pod instrumentation senses three-axis components of aircraft velocity, acceleration, attitude, and angular rate, as well as air data and weapon systems data. The AIS downlinks the tracking information to the TIS and relays pilot-initiated weapons firing signals to the CCS to initialize the simulation of weapon trajectories. Newer AIS pods and internal units are being designed to interface with the serial digital data bus of advanced aircraft (F-16/F-18 and beyond).

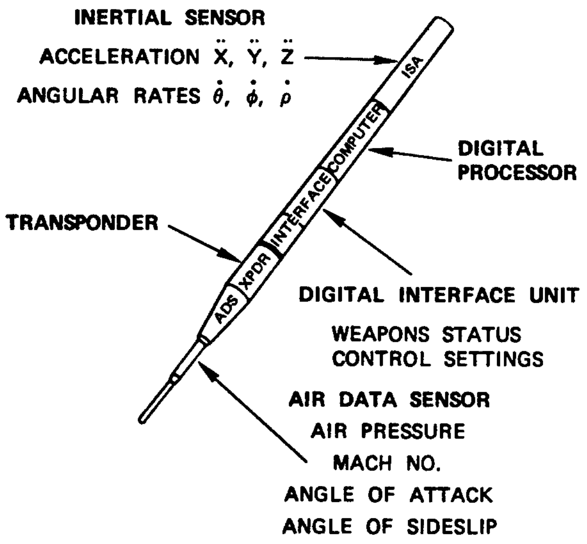
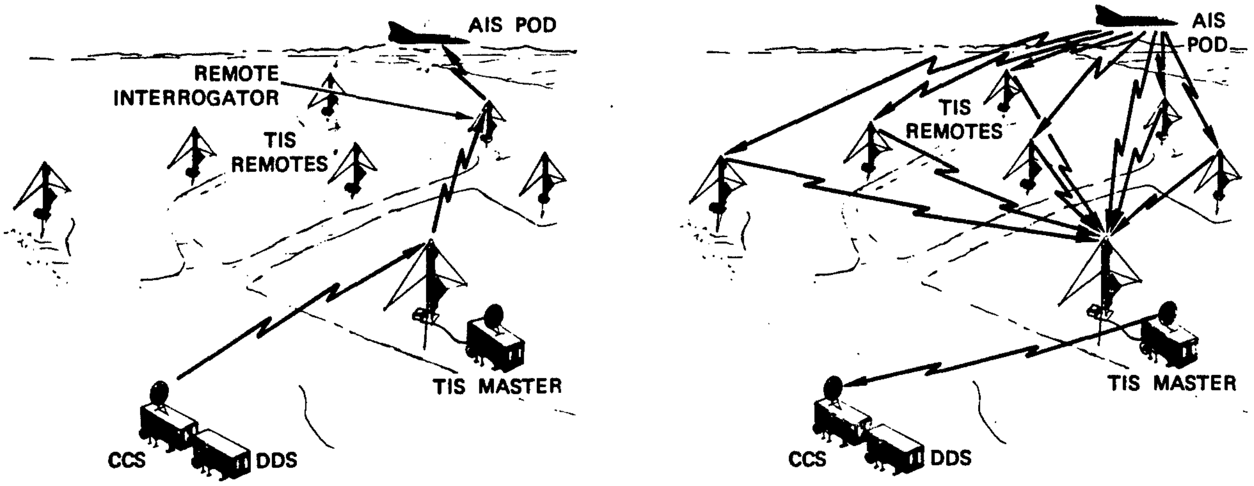


FIGURE 2 AIS POD SUBSYSTEM

2. Tracking Instrumentation Subsystem

Seven TIS ground stations and a master station are located to cover the exercise arena geometrically. The remote ground stations relay master-station transmissions to all AIS-equipped aircraft on the range and air-to-ground transmissions back to the master station. The master station contains a minicomputer and associated peripherals that control the TIS and process the raw range data. A full-duplex microwave data link allows communication between the TIS and the CCS. The master station may also be equipped with a ground-to-air-to-ground transmitter-receiver, so it can operate as the seventh ground station.

The TIS incorporates a frequency-modulated continuous wave phase-comparison ranging system for range measurements. Under computer control, this ranging system sequentially interrogates (Figure 3) each participant, using the remote sites to obtain multiple slant-range measurements to each aircraft during each interrogation. Although only three range measurements are required to determine position, the system attempts to collect all seven ranges to provide an overdetermined solution and to minimize data loss due to shadowing and aircraft blockage during maneuvers. The TIS relays raw range data and the downlink data messages to the CCS.



(a) INTERROGATION SEQUENCE

(b) RESPONSE SEQUENCE

FIGURE 3 AIS POD INTERROGATION AND RESPONSE SEQUENCES

3. Control and Computation Subsystem

The CCS, the "brain" of the TACTS/ACMI, consists of a large-capacity multiprocessor computational system and a microwave data link for communicating with the local and remote DDSs and with the TIS master station. Current systems use a multiprocessor configuration with shared memories.

A complex filtering technique is used to integrate the TIS-collected ranging and inertial data on up to 8 high-dynamic aircraft and to provide real-time state-vector tracking data to the DDS for display and recording. Ranging data are used in a separate filter for position-only tracking of 12 additional aircraft. Tracking results are available for display at the DDS within 0.2 s of real time. Using both the state-vector tracking information processed through the TIS and CCS computers and the timing of the weapons firing signals, the CCS calculates the miss distance between the simulated weapon and the target and scores the pilot accordingly.

Tracking data are also provided to a CCS subprogram that calculates flight safety. Aircraft are monitored for hazardous performance and for location within the range boundary. Appropriate alarms are transmitted to the RTO at the DDS when any selectable preset limit is violated.

To process an exercise, the CCS accepts exercise data for setup and termination from the DDS and responds to DDS-supplied fighter/target designations, hazard limit changes, and other exercise data. To sequence its programs correctly and to monitor pilot performance and hazards, the CCS compares fighter and target positions, velocities, and accelerations against predetermined parameters. In addition, the CCS maintains a statistical summary of results of all weapons used during each mission based on the weapon's launch-boundary compliance and kill determination data. The CCS performs real-time operability test and calibration and accepts status messages from the TIS and DDS, as well as from its own computer and peripheral equipment. These status data are included in the summary maintained by the CCS and communicated to the DDS.

4. Display and Debriefing Subsystem

Weapons results, aircraft maneuvers, and interaircraft parameters are graphically displayed in real time at the DDS facility (Figure 4) and are recorded for replay during postmission debriefing. Aircrews aloft also receive auditory notification of weapons results in real time. This capability teaches aircrews to recognize weapons-envelope boundaries and other parameters necessary for successful weapon delivery.

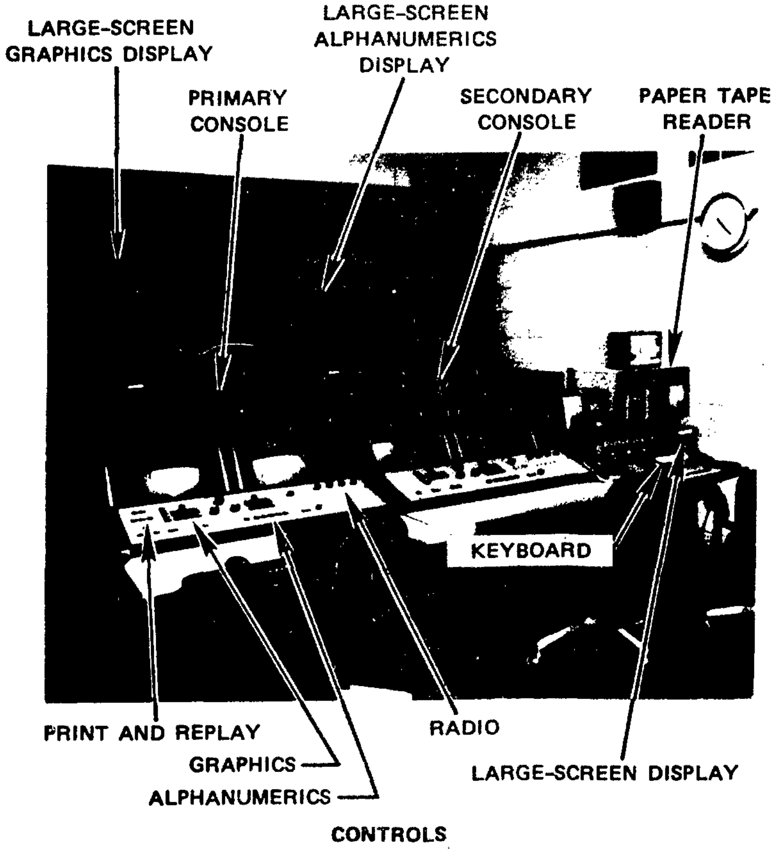


FIGURE 4 DDS DUAL-CONTROL DISPLAY CONSOLE

The DDS graphic, three-dimensional representation of the range and exercise activity uses perspective, varies size with distance, and reduces intensity with distance. Range terrain and aircraft locations, altitudes, attitudes, and flight paths are displayed on both situation-type CRT displays and large color observation screens. These situation displays have azimuth and elevation coordinate rotation, automatic centering of the exercise centroid, and "zoom" capabilities. Missions can thus be viewed from virtually any desired aspect in real time or during playback. Figure 5 shows examples of DDS situation displays. Maneuvering aircraft tracks are presented as variable moving images with selectable time-history ribbons. Aircraft ground tracks and range topography representation augment the three-dimensional character of the display. In addition, a pilot's view (Figure 5d) may be selected, which pictures the engagement from the cockpit of any selected aircraft on the range.

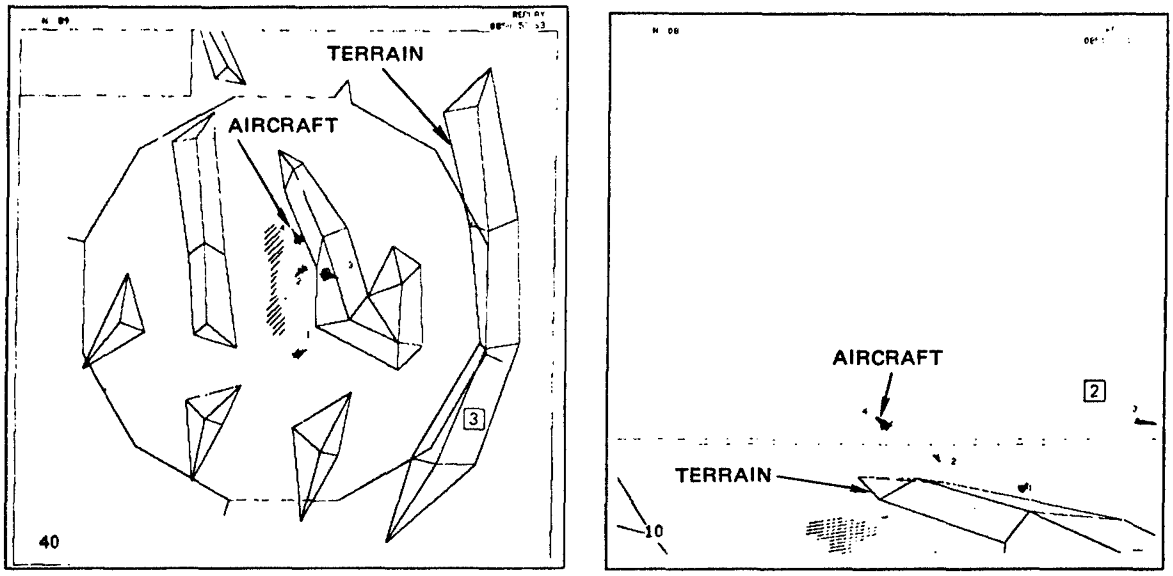
Dynamic flight data are displayed alphanumerically (Figure 6). Indicated airspeed, altitude, angle of attack and normal acceleration (g), as well as flight relationships between aircraft and target (range, bearing, closing velocity, angle of the tail, and antenna train angle), are typical of the information available in real time. Summary data are automatically displayed and printed at the end of each mission.

A selectable engineering data display (Figure 7) provides a wide range of engineering parameters, including the range position, velocity, acceleration, and attitude of each participant.

During replay, unlimited variations of the three-dimensional display can be selected for analysis, regardless of what was displayed during live monitoring. The operator has complete freedom to start, stop, change scale, and rotate coordinates. In addition, hard copies of any data or display can be printed for in-depth analysis. A more detailed technical description of TACTS is given by McHenry and York.[4]

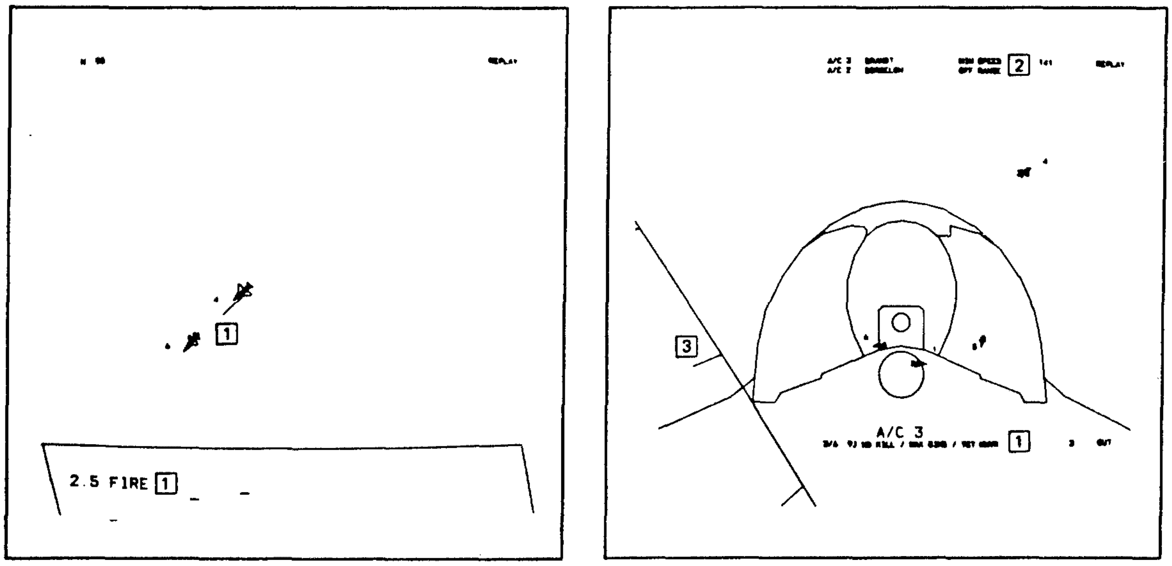
B. SRI's Ambulatory Physiologic Monitoring System (APMS)

SRI has also specialized in developing advanced physiological monitoring systems, particularly for use in an ambulatory environment, for many years. Typical projects include:



(a) PLAN VIEW (40-nmi scale)

(b) SLIGHT ELEVATION (10-nmi scale)



(c) MISSILE FIRE (2.5-nmi scale)

(d) PILOT VIEW (with reasons for miss)

FIGURE 5 DDS SITUATION DISPLAYS

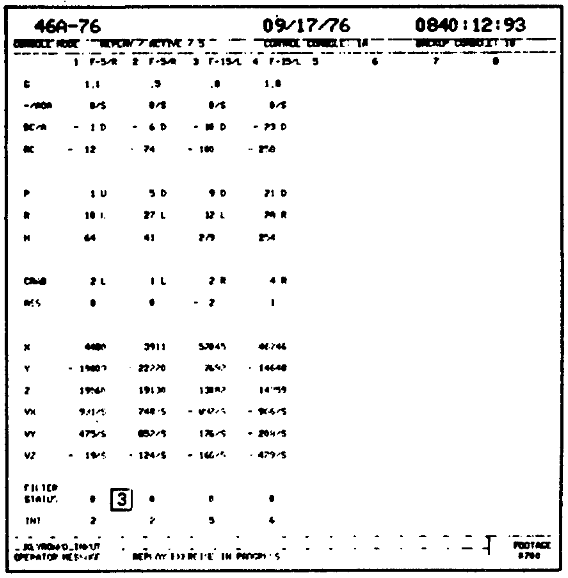
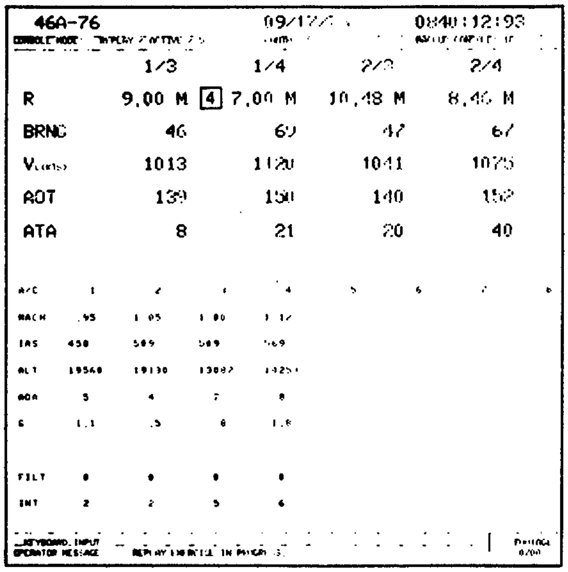


FIGURE 6 FLIGHT DATA DISPLAY



	1/3	1/4	2/3	2/4
R	9,00 M	$\boxed{4}$ 7,00 M	10,48 M	8,46 M
BRNG	46	69	47	67
$V_{(KTS)}$	1013	1120	1041	1075
AOT	139	150	140	152
ATA	8	21	20	40

FIGURE 7 DDS ENGINEERING DATA DISPLAY

- Ambulatory blood pressure monitor
- Ambulatory, 8-channel, 12-hour EEG recording system
- NASA Space Shuttle in-flight physiological monitoring system
- Stress test blood pressure monitor
- NASA venous occlusion cuff and controller.

Many of the instruments developed by SRI for NASA have been for use aboard the Space Shuttle. These instruments are built with the quality control and environmental testing required for flight certification.

Several systems have been combined to form the Ambulatory Physiologic Monitoring System (APMS), the elements of which are described below.

1. Recording System

A key element in many of SRI's ambulatory monitoring systems is the 8-channel analog or digital cassette recording system. This recording system has been developed over the past nine years under the auspices of (first) the National Institutes of Health, for long-term EEG recording, and (presently) NASA, as a general-purpose data recorder. The system, which evolved through several versions of both recorder and playback systems, is now fully operational and has been certified for NASA use in Space Shuttle programs.

The current recorder system consists of a battery-powered, single-unit, digital cassette recorder (Figure 8) and a desk-top, microprocessor-based, data playback system (Figure 9). The recorder has eight independent data channels. These channels can be configured to be all digital, or six digital and two analog channels. The digital channels are digitized to 10 bits, giving a 0.1% accuracy and 60-dB dynamic range. The analog channels, if used, have a 2% accuracy and 40-dB dynamic range. The analog channels are capable, however, of a higher frequency response than the digital channels. The recorder can be configured for three different recording speeds, allowing recording time to vary between 2 and 10 hours. Available bandwidths and recording times are given in Table 1. A time code can be inserted once a minute on the digital channels, does not require a separate data channel, and can be used to synchronize the recorder to the TACTS system. The recorder gains in flexibility, small size, and low power consumption from the use of complementary metal-oxide-semiconductor (CMOS) microprocessor technology packaged in custom hybrid modules.

2. Playback System

A playback system is used to recover the data from the cassette recorder. Data are played back faster than real time and are transferred directly to a medium-sized computer. To do this, a bipolar bit slice microprocessor controls data playback and data transfer. Once transferred and stored, the data can be examined and processed at will using the capabilities of the host computer.

The primary advantages of this recorder system are that: (1) eight channels of data can be recorded, (2) digital recording provides signal fidelity equivalent to laboratory conditions (accuracy of 0.1%),

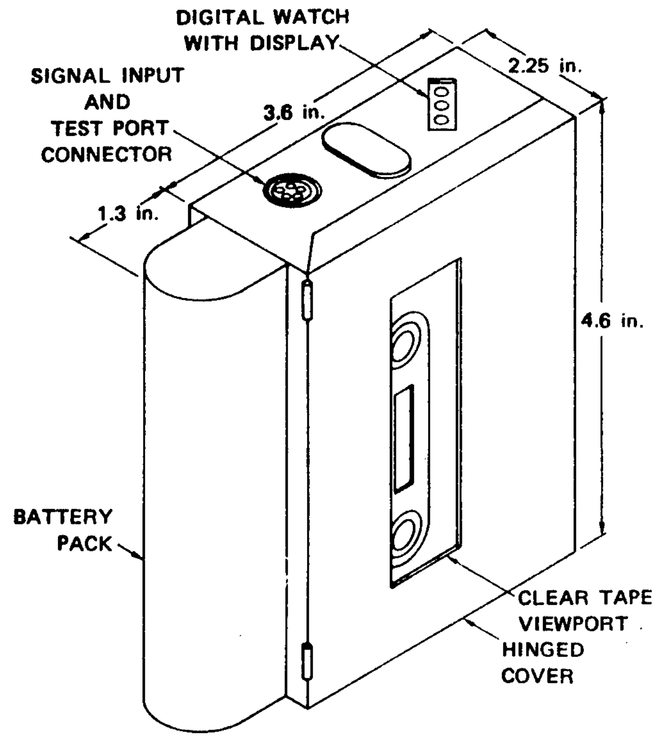


FIGURE 8 EXISTING 8-CHANNEL EEG RECORDER, ELECTRONICS, AND BATTERY PACK

(3) long-term recording is available, (4) the recorder is small in size, and (5) data playback and processing is convenient and flexible.

3. Ambulatory Blood Pressure System

For SRI's APMS, the cassette data recorder is combined with cuff inflation and control and with appropriate sensors and signal conditioning to create a system that records all the data necessary to determine blood pressure, even when the subject is moving. A motor and air pump combination inflates the cuff. Two miniature air valves under microprocessor control gradually deflate and empty the cuff as desired. Two pressure transducers provide backup system control and cuff pressure data. A microphone is used to obtain Korotkov sounds, and ECG electrodes are used to obtain high-quality ECG data for recording. The system is powered from rechargeable nickel-cadmium batteries. In addition to the armcuff, the system is configured in two belt-worn packages, one containing the battery pack, motor, air pump, and pneumatics, and the other

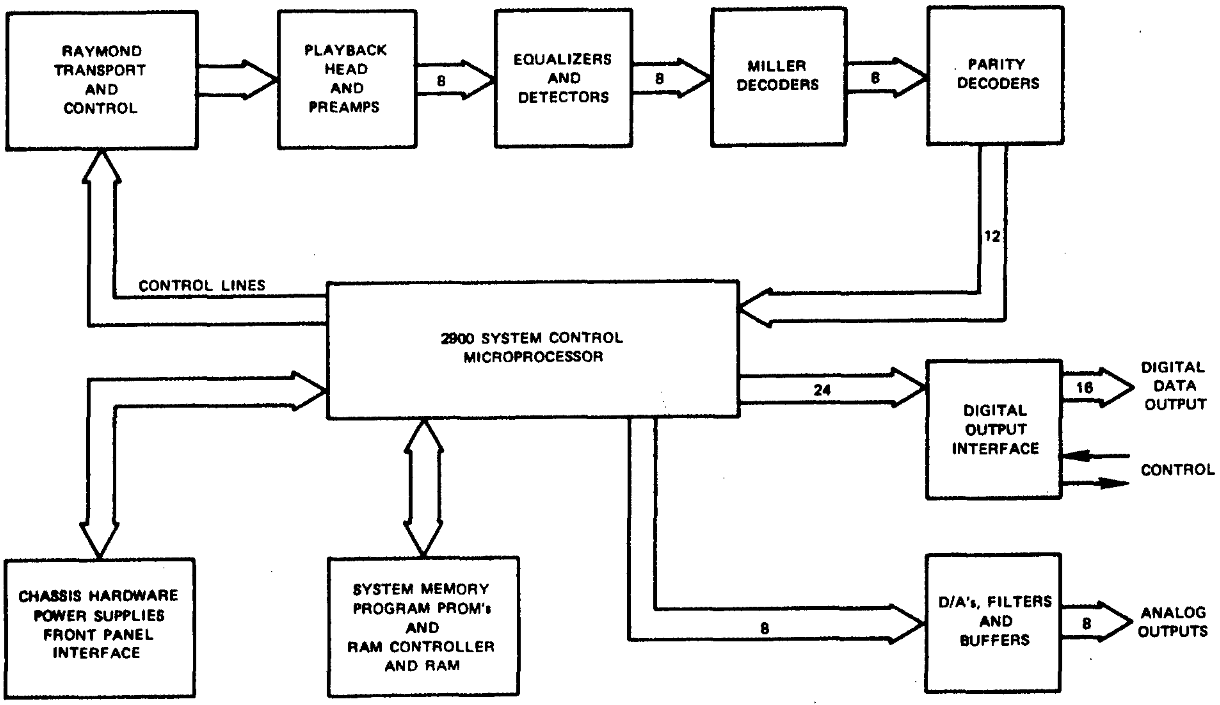


FIGURE 9 EXISTING PLAYBACK SYSTEM (block diagram)

Table 1

PROPOSED CASSETTE DATA RECORDER PERFORMANCE CHARACTERISTICS

Tape Speed (in/s)		Digital Sample Rate (samples/s/channel)	Digital Channel Encoded Analog Bandwidth (Hz)	Digital Data (bits/s/channel)	Data Density (bits/in./track)	Tape Record Time (h)		Analog Bandwidth per Channel (kHz)	Analog Dynamic Range (dB)
Dec.	Frac.					C-90	C-120		
0.1875	3/16	100	40	1100	5866	8	10.6	1.5	30
0.375	3/8	200	80	2200	5866	4	5.3	3.0	35
0.75	3/4	400	160	4400	5866	2	2.6	6.0	40

containing the recorder, cuff control system, and user interface and controls.

This ambulatory blood pressure system features (1) preprogrammed timed intervals between inflations, (2) preprogrammed pressure limits, (3) user-initiated inflation or deflation override, (4) deflation rate based on heart rate, (5) user display showing cuff pressure during inflation, and (6) numerous backup safety provisions to protect against overinflation and excessive-time inflation.

The system can inflate to pressures up to 300 mmHg and can run continuously for 4 hours with up to 100 inflations. Longer monitoring times are available if intermittent recording is provided.

Blood pressure data are obtained after the data cassette is retrieved and played back through the playback system. A highly sophisticated digital filtering and processing algorithm is used to examine the recovered Korotkov sounds, ECG signal, and cuff pressure to determine blood pressure. This algorithm can obtain accurate blood pressures even in the presence of high external noise levels and subject motion artifacts. In fact, this system has been used successfully in obtaining automated blood pressure measurements during both bicycle and treadmill stress testing.

4. Antigravity Suit Evaluation Instrumentation

A third system, the antigravity suit in-flight evaluation instrumentation, developed by SRI for use by NASA on the Space Shuttle, combines the cassette data recorder and the blood pressure measuring system with Doppler blood flow instrumentation, phonocardiogram, and g-suit pressure monitors to form a highly sophisticated monitoring system. The system is battery-powered and self-contained in a single package that attaches to the seat of the individual being monitored. All data, including the raw data to determine Doppler blood flow, are recorded. During data playback, the various blood pressure and blood flow data are processed and recovered, along with the ECG, phonocardiogram, and antigravity suit pressure waveforms.

SRI has also built the hardware necessary to measure body temperature and respiration, but has not yet integrated it into the APMS. Temperature is measured with a small semiconductor temperature probe; respiration is measured with SRI's piezoelectric chest expansion strain guage. These potential additions to the APMS are described in Section IV.

III CURRENT DATA MEASUREMENT CAPABILITIES

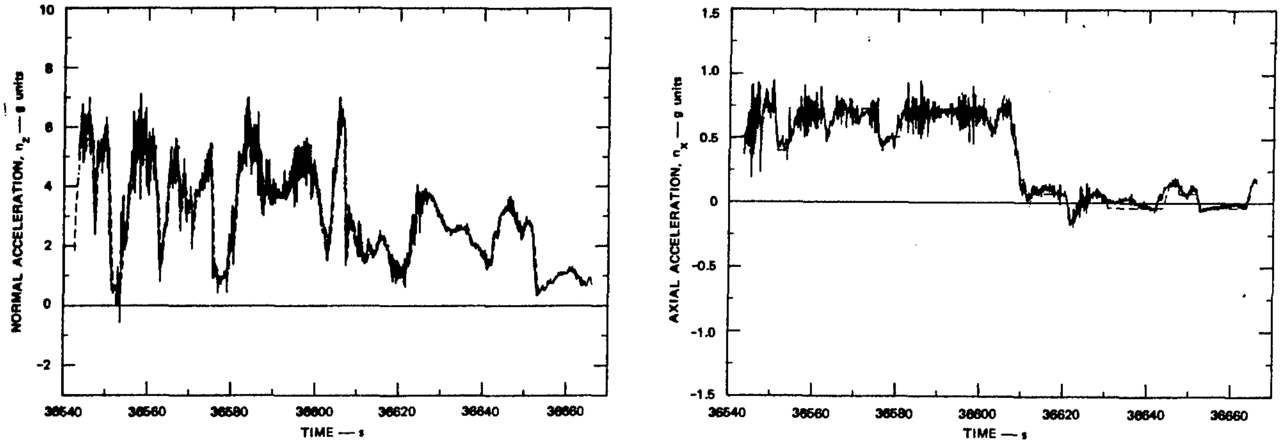
Pilot performance depends on optimal weapons selection, evaluation of the tactical situation, and accurate targeting, as well as on aircraft and weapon performance. Likewise, the overall effectiveness of the aircraft and the weapon system depends on the pilot's ability to operate them. We therefore need to determine the optimal level of stress at which the pilot is most alert, yet able to accommodate the many decisions facing him. This section describes SRI's current capabilities for measuring both aircraft performance and pilot stress and workload.

A. Aircraft Data

The TACTS/ACMI system provides complete aircraft state-vector tracking data, including load factors, velocities, position, angular rates, attitudes, and air data sensor information. Table 2 lists the specific data available from the TACTS/ACMI system. Examples of these data are presented graphically in Figures 10-12, which show both TACTS/ACMI and onboard reference instrumentation data. The TACTS/ACMI data were taken directly from CCS range tapes. In all cases, the range data compare closely to the aircraft reference data. Figure 10 shows examples of state-vector tracking data. Figure 11 shows examples of air data sensor information available from the range. Finally, Figure 12 is an example of a parameter derived from the data provided by the system.

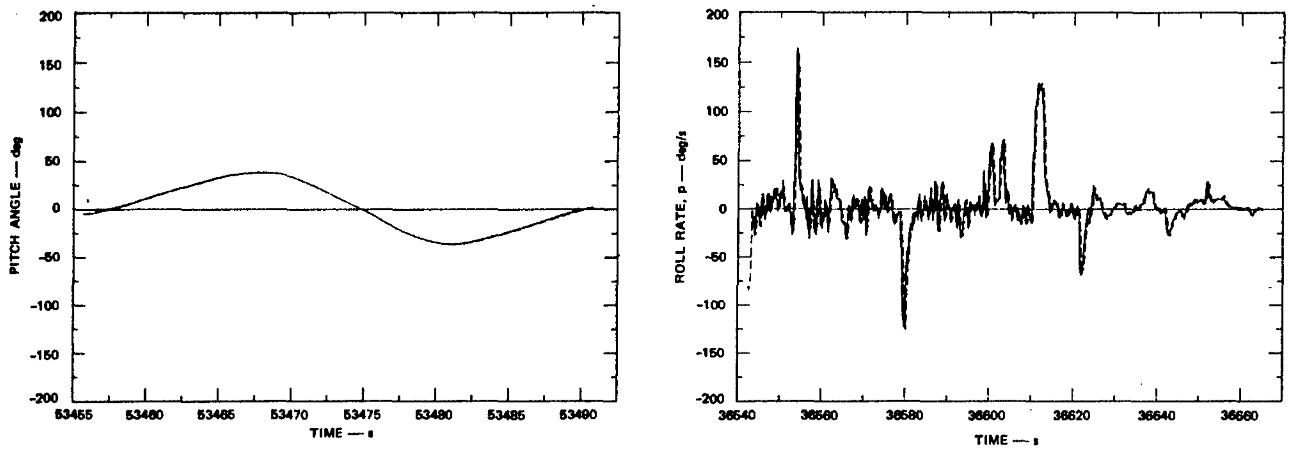
B. Pilot Physiologic Data

The current SRI APMS can be used conveniently to obtain and record biologic signals of respiration, heart rate, and blood pressure from a subject under operational conditions. To obtain these measures, several transducers are placed on the test subject: one ground electrode, two ECG electrodes, one solid-state chest strain gauge, and a standard blood pressure armcuff. A small microphone embedded in the armcuff records the Korotkov sounds during the periodic blood pressure measurements.



(a) NORMAL ACCELERATION AS A FUNCTION OF TIME

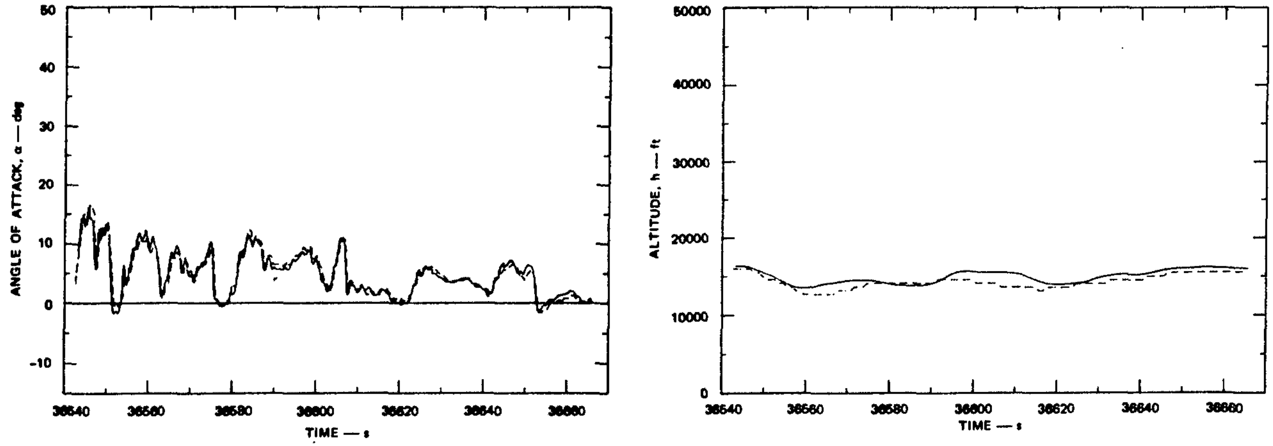
(b) AXIAL ACCELERATION AS A FUNCTION OF TIME



(c) PITCH ANGLE AS A FUNCTION OF TIME (Yuma event)

(d) ROLL ANGLE AS A FUNCTION OF TIME (Nellis event)

FIGURE 10 TACTS STATE-VECTOR TRACKING DATA



(a) ADJUSTED ANGLE OF ATTACH AS A FUNCTION OF TIME (Nellis event)

(b) ATTITUDE AS A FUNCTION OF TIME (Nellis event)

FIGURE 11 TACTS AIR DATA SENSORY INFORMATION

Table 2

TACTS/ACMI DATA CAPABILITY

State Vector Tracking Data

- Position (x, y, z)
- Velocity (x, y, z)
- Acceleration (x, y, z)
- Attitude (w_x, w_y, w_z)

Derived Data

- Angular acceleration
- Weight
- Altitude (above terrain)

Air Data Sensor Information

- Dynamic pressure
- Static pressure
- Mach number
- Calibrated airspeed
- Angle of attack
- Angle of sideslip

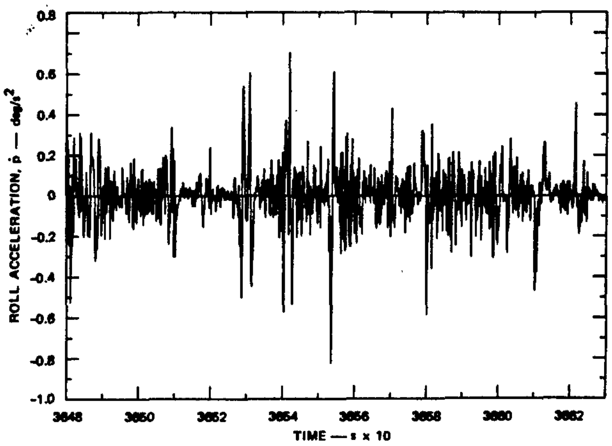


FIGURE 12 DERIVED DATA (roll acceleration)

1. Respiration

Respiratory signals can be acquired by a variety of transducers. The APMS uses an SRI-developed piezoelectric chest-wall strain gauge to monitor respiration without the restrictions of ordinary gauges. Mercury-tube strain gauges, for example, though commonly used for sensing chest and abdomen expansion, require a sophisticated amplifier. A more important disadvantage is that mercury is quite toxic, and the risk of a broken gauge is great. Impedance pneumography is subject to body movement and heart motion artifact. Pneumotachography gives good data, but restricts the subject's movements.

The SRI strain gauge has none of these disadvantages. Figure 13 shows a prototype version. The transducer is a piezoelectric crystal about the size of a matchhead. Connections to the crystal and its rubber housing combine to produce a package about the size of the eraser on the end of a pencil. The preamplifier for the unit requires only two integrated circuits. The output from the preamplifier is a high-level (i.e., hundreds of millivolts) voltage directly proportional to the strain exerted on the transducer.

In operation, the transducer is placed in the middle of the back. Its strings are cinched lightly around the torso with a double D-ring. The cinching should be tight enough to ensure that the transducer is always in tension, but not tight enough to restrict breathing. A little elasticity in the strings ensures the latter condition.

Figure 14 shows the output of the SRI strain gauge compared with a reference mercury-filled strain gauge. For this comparison, the SRI gauge girdled the entire abdomen, while the mercury device was sensing motion only on one side of the abdomen. The device provides respiration signals independent of body position and reasonably free of motion artifact.

2. Heart Rate

SRI's ambulatory recorder systems can record ECG signals accurately (to 0.1%). Upon data playback, this ECG signal can be analyzed to

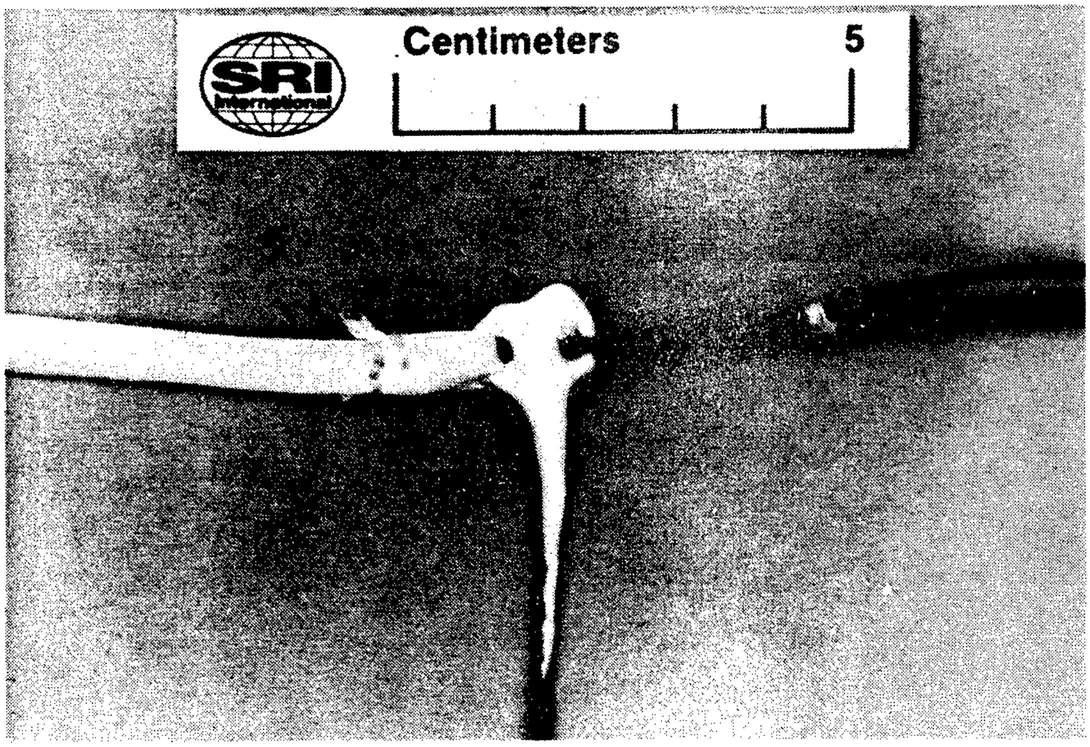


FIGURE 13 PROTOTYPE SRI STRAIN GAUGE

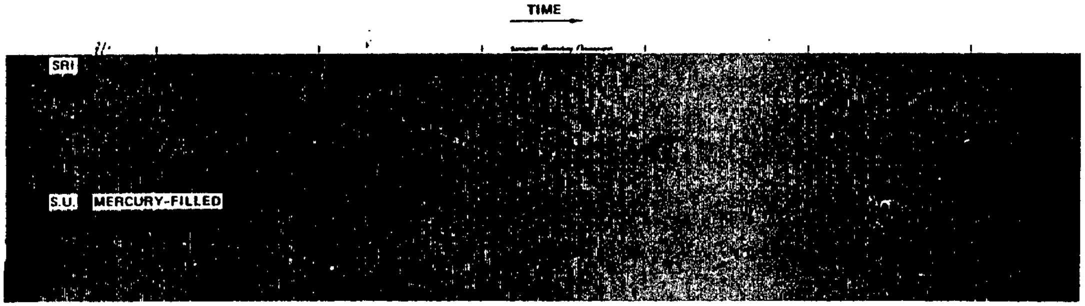


FIGURE 14 SRI STRAIN GAUGE COMPARED WITH MERCURY-FILLED STRAIN GAUGE

determine R-to-R intervals, average heart beat, and (because of the recording fidelity) S-T segment depression. Other ECG data analysis techniques could also be used, if desired.

3. Blood Pressure

Of the three physiologic parameters measured by the APMS, blood pressure is by far the most difficult to obtain. SRI has designed and constructed several ambulatory blood pressure measuring systems and currently has an operating prototype of the APMS system described in this paper.

After the data have been obtained using the on-body portion of the APMS (Section II), the recorded data are played back through the cassette data-playback system and transferred to a computer. The computer is then used to analyze the data and to determine systolic and diastolic blood pressure.

The software package uses the intervals between the ECG R-wave peaks and the onset of the Korotkov sounds (the R-K interval) during a cuff deflation to determine the systolic and diastolic blood pressures each time the blood pressure cuff is inflated. A sophisticated computer algorithm is used to determine systolic and diastolic blood pressures. This algorithm (a detailed description of which is beyond the scope of this paper) makes it possible to obtain accurate blood pressure readings in the presence of external noise and subject movement (even strenuous physical activity) that normally would mask the Korotkov sounds and make accurate blood pressure readings impossible.

C. Pilot Tasking Data

As part of the TACTS/ACMI system, all weapons selection data are recorded during real-time combat exercises and are available for later replay to determine whether the pilot best utilized the armament available. Weapons data currently recorded on the TACTS/ACMI system include: bombs, mines, Falcon (AIM-4), Sparrow (AIM-7), Shrike (AIM-45), and Sidewinder (AIM-9). The individual signals available from each weapon are listed in Table 3.

Table 3

INDIVIDUAL SIGNALS ASSOCIATED WITH SPECIFIC WEAPON TYPE

Bombs

Trigger

Mines

Trigger

Shrike (AIM-54)

Fire ready
Fire
Direction finding angles
Pulse repetition frequency

Sidewinder (AIM-9)

Coolant
Seeker angles
Trigger-arm
Audio

Falcon (AIM-4)

Trigger-arm
Heat
Radar

Gun select
Trigger-safe

Sparrow (AIM-7)

Head aim angles
Range
Trigger
Gate select
CW power
Dogfight

IV FUTURE DATA MEASUREMENT CAPABILITIES

A. Aircraft Data

Plans are underway for an extensive TACTS/ACMI evaluation of both the F-16 and F/A-18 aircraft. For these operations, the TACTS/ACMI is being adapted to collect aircraft performance data from the data multiplex bus in these new aircraft. The additional data that will be available will expand the TACTS/ACMI utility for both the training and test communities of users. The TACTS/ACMI system will be able to access any aircraft data that are available on the avionics subsystem. For example, if engine performance data are required, the avionics can supply rpm, inlet temperature, outlet temperature, fuel flow, and power lever angle. If specific data are needed to analyze flight dynamics, the serial data capability can supply flight control surface position and rates. The future data accessability of the TACTS/ACMI system will increase significantly as new and modified aircraft begin training with this system.

B. Pilot Physiological Data

SRI plans to expand the present APMS capabilities to measure skin or rectal temperature, temporal blood flow, peripheral blood flow and pooling, and cardiac output. The technology required for the ambulatory measurement of these additional parameters is fully developed and has been demonstrated on centrifuge tests at NASA-Ames Research Center.

1. Temperature

SRI uses a semiconductor transducer to measure body temperature. This temperature-measuring circuit may be used to obtain skin, rectal, or tympanic membrane temperature.

2. Temporal Blood Flow

In the past, human $+G_z$ acceleration tolerance has partly relied on subjective criteria related to peripheral loss of vision. A noninvasive Doppler flow system has recently been developed by Dr. Salvadore A. Rositano at NASA-Ames Research Center. This system may be used to obtain an objective end-point criteria for $+G_z$ tolerance. It consists of miniature 8-MHz Doppler sensors placed on the forehead over both frontal branches of the temporal arteries to detect blood flow to the head from backscattered ultrasound. As reported by Dr. Rositano,[1]

> Over 100 subjects have now been studied during more than 2,000 centrifuge runs. Objective changes in temporal artery flow velocity consistently preceded visual degradation for each subject during all acceleration profiles. No subject has gone unconscious without first exhibiting a minimum 6 seconds of total flow cessation. Retrograde flow followed by complete flow cessation always preceded central light loss.
>
> Results indicate that this method can be successfully used with a wide variety of tasks during exposure to $+G_z$ acceleration. It is recommended for use during evaluation of protective maneuvers or devices on the centrifuge or during actual flight in high performance aircraft.

A typical data trace obtained during Rapid Onset Run (ROR) $+G_z$ centrifugation is shown in Figure 15. As stated in this report,

> When blackout was approached (range 2.7 to 4.6 G), eye level arterial blood pressure (Pa) began to fall concommitant with the occurrence of retrograde flow in the temporal artery. Zero forward temporal artery flow (Qta) was determined by both graphic and audio recordings six seconds (range 4-9 s) prior to blackout. Eye level mean arterial pressure (Pa) decreased to 20 ± 1 mmHg ($n = 16$) when zero forward Qta was initially recorded. Arterial pressure and temporal artery flow velocity increased simultaneously during centrifuge deceleration, with a significant increase in pressure and flow velocity occurring post run when compared to previous values.

The point of blackout is denoted by the vertical line in Figure 15, where the subject's central light loss (CLL) occurred.

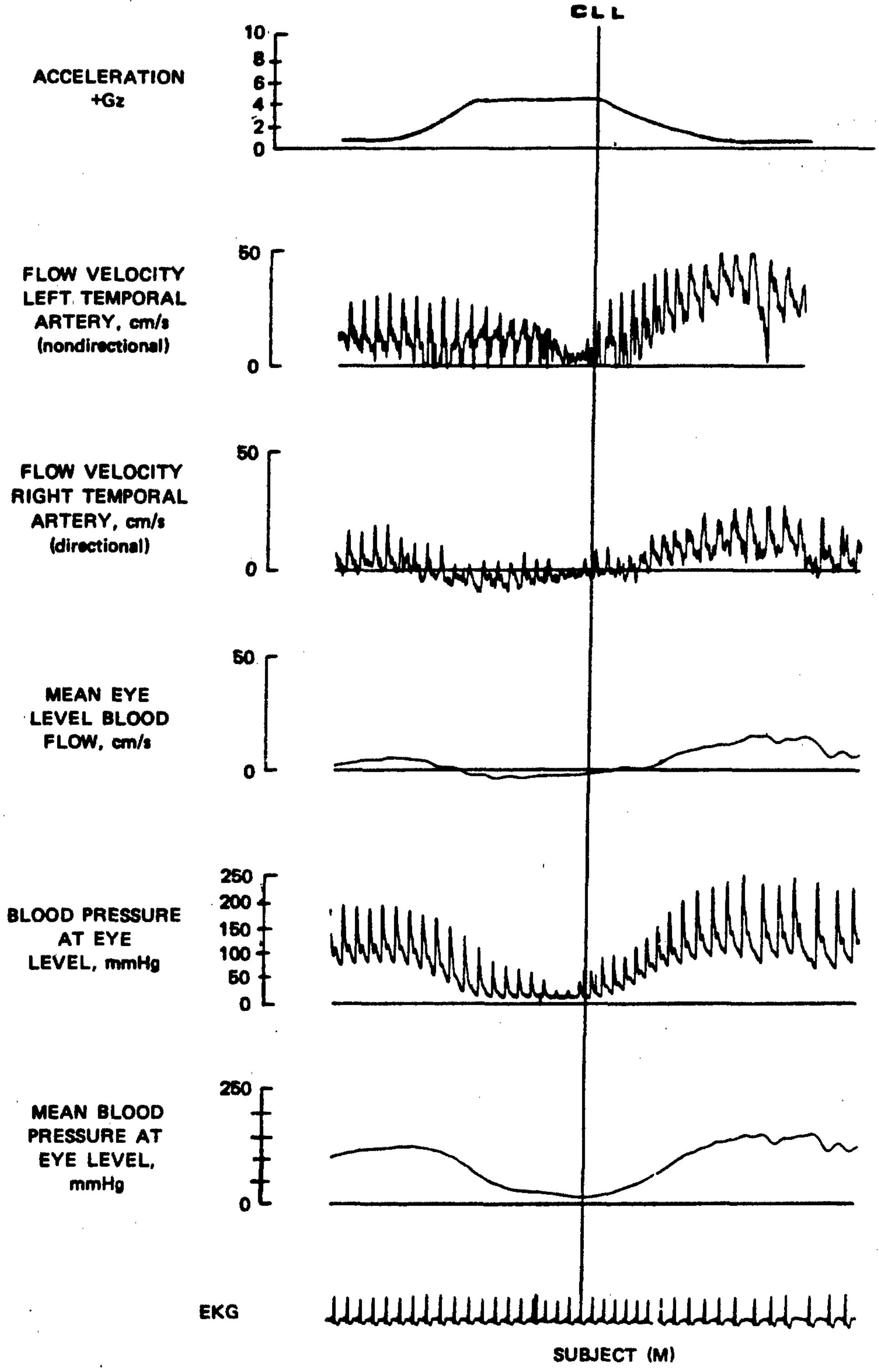


FIGURE 15 EFFECT OF RAPID ONSET ACCELERATION ON EYE-LEVEL BLOOD PRESSURE AND FLOW VELOCITY

Rositano's system is currently being modified for use during Space Shuttle flights. It will be incorporated into a flight instrument package being built by SRI for NASA-Johnson Space Center. The output signals from the Doppler sensors will be recorded on the SRI cassette tape recorder as used in the APMS.

3. Peripheral Blood Flow

Peripheral blood flow and pooling have also been measured during $+G_z$ centrifugation using impedance plethysmography. This work was done during the past five years as part of male/female bed-rest studies conducted by NASA at the Ames Research Center. This technique was used to quantify the hemodynamic responses of 45- to 65-year-old subjects to Space Shuttle reentry $+G_z$ acceleration before and after a 10-day simulated exposure to weightlessness. These measures were also used to assess the effectiveness of wearing an inflated antigravity flight suit during simulated reentry acceleration.

Each subject was instrumented as shown in Figure 16. The impedance plethysmograph introduced a high-frequency (about 50 kHz), low-amperage (about 0.1 mA rms) electrical current between shoulder and foot disposable EKG electrodes (I_1 and I_2). Simultaneous measurements of baseline resistance (R_o) and pulse volume resistance changes (R) were made in the leg between similar ankle and upper-thigh recording electrodes. Records of R_o were analyzed according to Nyboer[2] to determine the amount of blood accumulated in the leg during each test sequence. Leg blood flow indices were calculated using the segmental R values.[3]

The acceleration tests took place during bedrest studies that consisted of a 9-day control period, 10 days of absolute bed rest, and 5 days of recovery. Each subject was exposed to $+G_z$ centrifugation six times during the study—on days 1, 2, and 7 of the control period to establish a baseline for acceleration tolerance, and on bed-rest days 7 through 9 to determine the incidence of deconditioning. During all centrifugation exposures, the subjects were tested both with and without inflated antigravity suits to evaluate the effectiveness of this

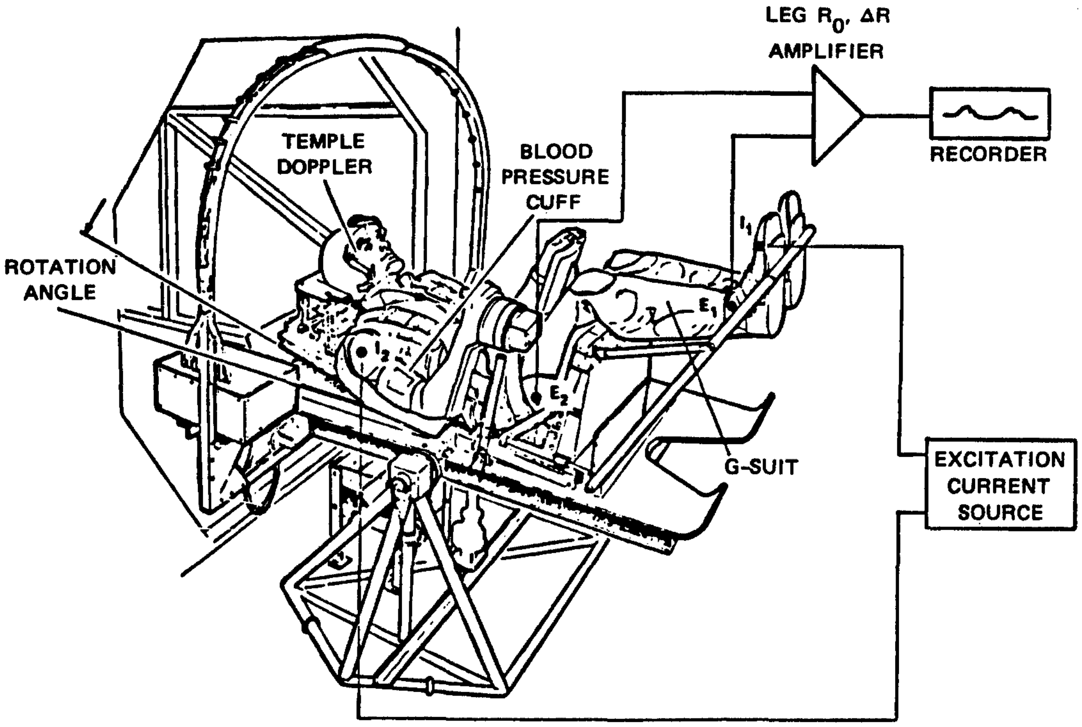


FIGURE 16 INSTRUMENTED SUBJECT DURING +Gz CENTRIFUGE TESTS

countermeasure. On days 1 and 2 of the control period and days 7 and 8 of bed rest, the subjects were exposed to acceleration profiles of 1.5 $+G_z$ and 2.0 $+G_z$--the levels projected for Space Shuttle reentry (see Figure 17). Each of these tests included 10 min in the upright position to simulate 1.0 $+G_z$ during landing following the simulated reentry acceleration stress. On day 3 of the control period and on day 9 of bed rest, the subjects underwent a +3.0 G_z acceleration exposure as shown in Figure 17 for 15 min as a general test of acceleration tolerance. During each suited test, the antigravity suits were inflated 1.0 psia/g.

Figure 18 shows a typical data trace obtained during one of the +3.0 G_z tests before bed rest. The antigravity suit was not inflated during this test sequence. The top trace in each portion of Figure 18 shows the subject's ECG signal. The bottom trace in each segment shows the subject's leg blood flow pulse. Figure 18a was taken at 2-min elapsed time during the resting control period preceding centrifugation.

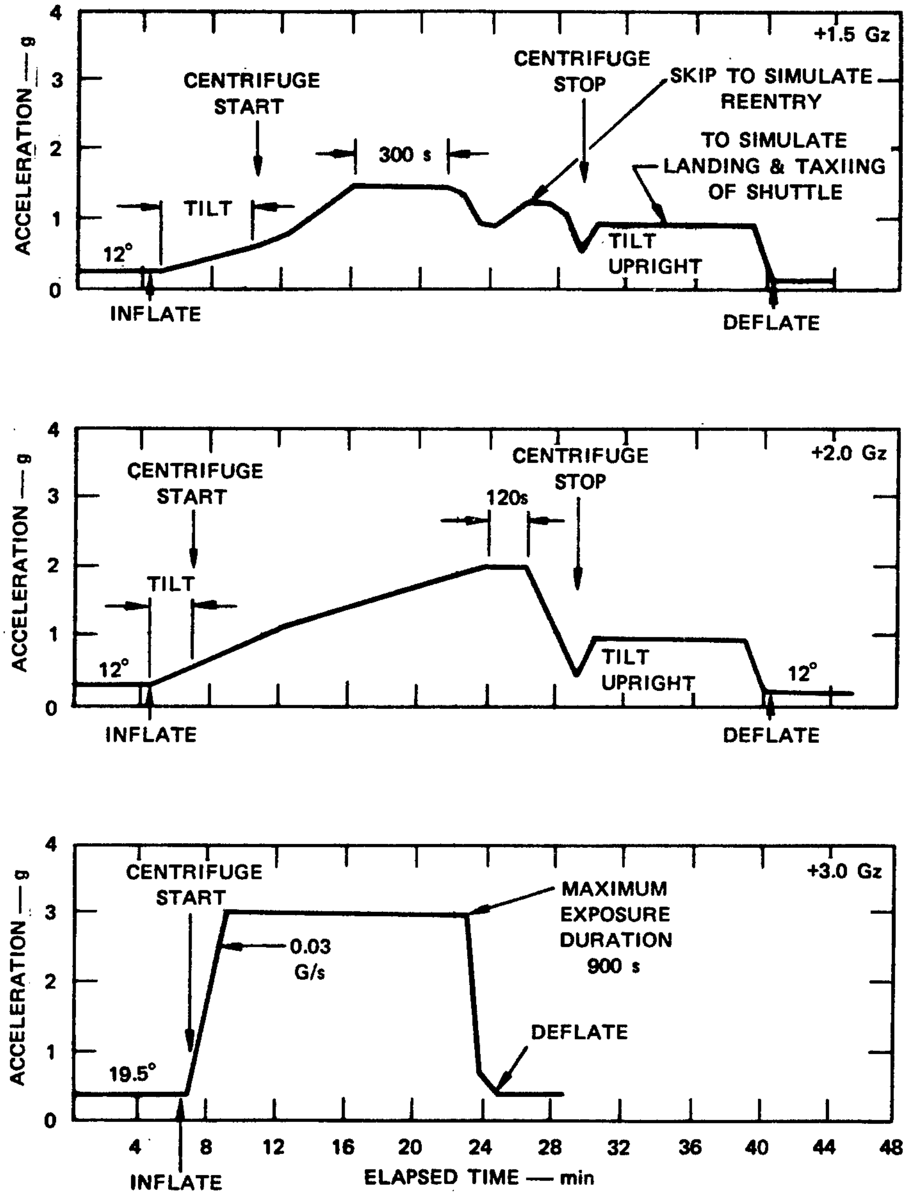
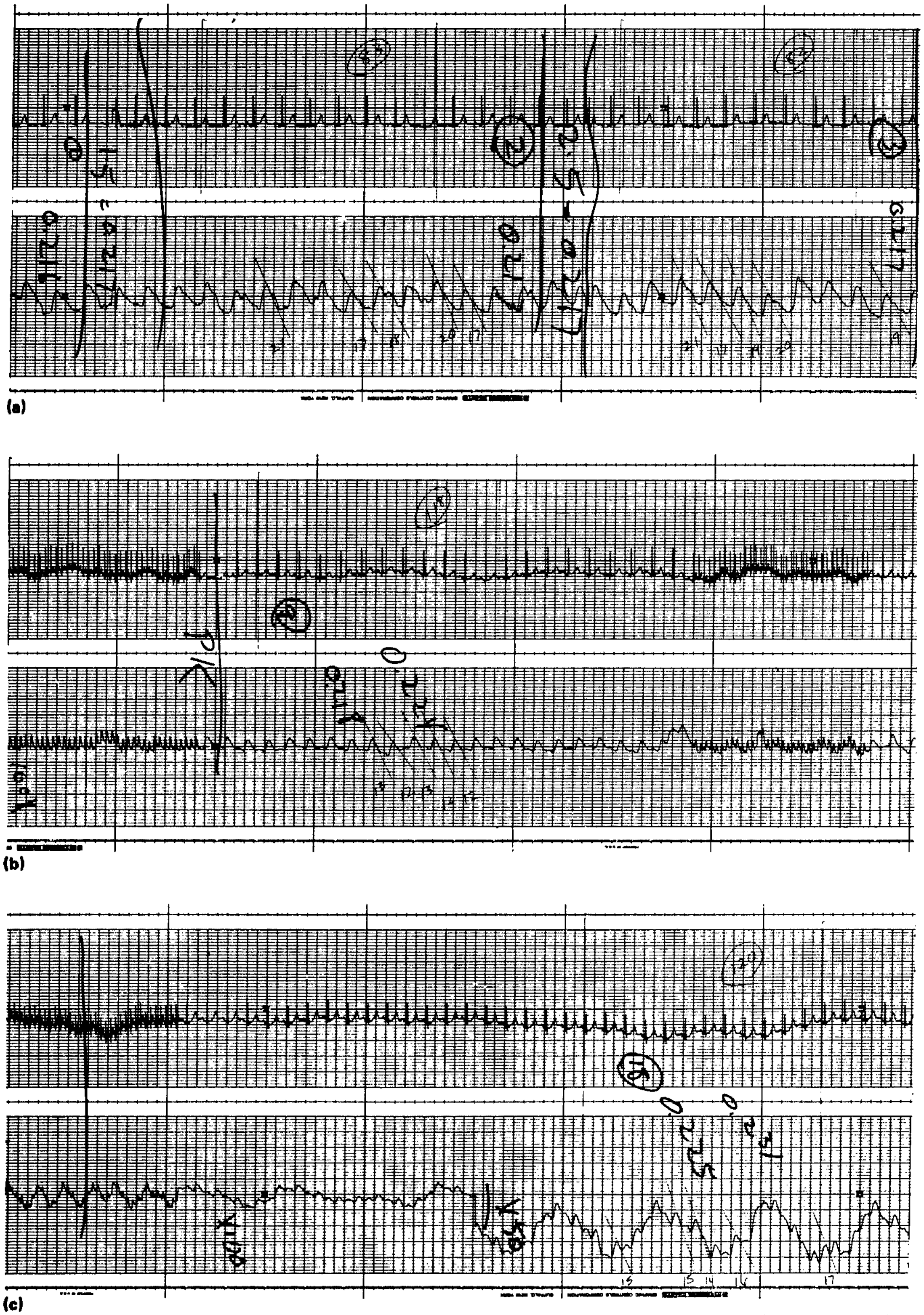


FIGURE 17 ACCELERATION PROFILES USED DURING CENTRIFUGE TESTS



SUBJECT 78 — 2/11/78
+3.0 G_z — PRE

FIGURE 18 IMPEDANCE DATA TRACE FROM CENTRIFUGE TESTS

Figure 18b shows the same data traces obtained at the time that the subject initially reached +3.0 G_z acceleration at 9-min elapsed time. Figure 18c shows the traces obtained after 9-min exposure to +3.0 G_z (at 18-min elapsed time).

By comparing the pulse amplitudes the R traces at the various times (use the X100 sensitivity at 18 min in Figure 18), one can determine the extent of decreased perfusion that is produced by increased $+G_z$ stress.

The grouped mean (±S.E.), percent leg pooling (% DVE) and leg blood flow (% LBF) values obtained during the 1.5 $+G_z$ pre-bed-rest tests of 45- to 55-year-old men are shown in Figure 19. These data illustrate the hemodynamic responses to the Space Shuttle reentry profile (Figure 19a) and show how these measures are altered by inflation of the antigravity flight suit. The asterisks in Figure 19 denote significant (P 0.05) changes that occurred for each parameter between the various important events/times during the $+G_z$ profile. The plus symbols show significant (P 0.05) differences between the inflated (dashed lines) and the non-inflated (solid lines) tests at specific times.

These results demonstrate that impedance plethysmography can be used to quantify hemodynamic responses of flight-suited subjects to periods of dynamic stress. These post studies used a rack-mounted impedance unit. This equipment could be miniaturized, however, and incorporated into the APMS instrumentation for ambulatory measurements.

4. Cardiac Output

Impedance plethysmography has also been used by others[4,5] to measure cardiac output and stroke volume during exercise. An ambulatory impedance instrument would extend this capability into the high-performance aircraft environment.

C. Pilot Tasking Data

Table 4 lists the cockpit controls and displays that will be monitored when the TACTS/ACMI system is used for F/A-18 training. Table 4

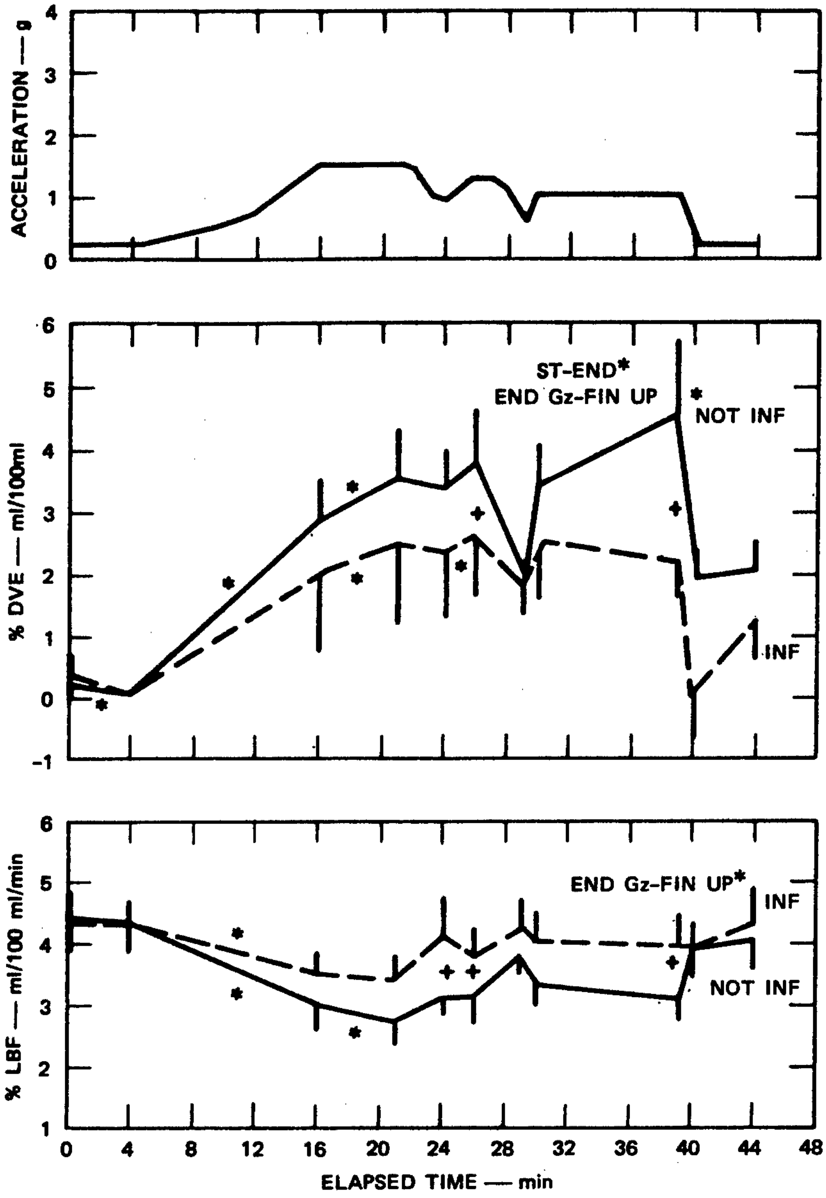


FIGURE 19 GROUP MEAN ± STANDARD ERROR BY HEMODYNAMIC RESPONSES TO 1.5-g REENTRY PROFILE

does not list each discrete signal associated with a given display, control, or weapon, but it does give the general description of its function. Each function normally has several selectable discretes associated with it. When the system becomes operational, each discrete will be available through the TACTS/ACMI system. The system will be capable of duplicating in the replay mode all tasks associated with display utilization, control inputs, and weapon selection.

Table 4

F/A-18 COCKPIT CONTROLS AND DISPLAYS

Head-Up Display

- Basic flight data
- Flight path/pitch ladder
- Waterline symbol
- Velocity vector caging
- Clutter reject
- Landing and steering data
- Advisory data

Flight Control Stick and Throttle Switches

- A/A weapon select
- A/G weapon release
- Gun missile trigger
- Flight control
- Display select
- Radar mode select
- Communication select
- EW select

Air-to-Air Mode

- Gun
- Sparrow (AIM-7F)
- Sidewinder (AIM-9L)
- Radar control

Air-to-Ground Mode

- Conventional bombs
- Laser-guided bombs
- Rockets
- Maverick (AGM-65E)
- HARM (AGM-88A)
- Gun

Electronic Warfare Displays

- Chaff
- Flare
- Jammer
- Dispenser
- ECM
- ECCM
- FLIR
- LST
- Radar mode selection
- Radar terrain avoidance mode

V SUMMARY

A. Recorded Data Advantages

The use of a digital data acquisition system significantly improves the quality of recorded measurements. The primary advantages of this technology include: (1) higher accuracy of recording, (2) increased number of channels available, and (3) greater dynamic range. This permits more sophisticated analysis of trace waveforms than would be possible with analog data acquisition systems. Subtle changes in waveform morphology can be detected and analyzed. In addition for physiological measurements, pattern recognition and artifact rejection can easily be accomplished during post-flight reduction of the recorded data.

B. System Performance Assessment

Today's technology can for the first time provide an operational assessment of the complete airborne weapon system, including the aircraft, pilot, and armament. It is possible to describe aircraft performance, pilot performance, and weapon performance quantitatively and to relate these figures to mission success.

Figure 20 illustrates the various links that associate mission success with the three major system components (aircraft, pilot, and armament). For example, if a mission were not successful but the aircraft performed satisfactorily, the problem may be with the pilot. Further investigation may show that blood flow to the brain was reduced during a high-g maneuver, thus limiting the pilot's ability to fly the mission properly. Or it may simply be that a given aircraft has superior performance (i.e., speed or turn-rate advantage) over a given adversary. The system described in this paper is a relatively inexpensive tool capable of statistically identifying the individual parameters that limit system performance, whether they are associated with the aircraft, the pilot, or the armament system. Summary statistics could be generated to identify consistent problem areas associated with poor performance.

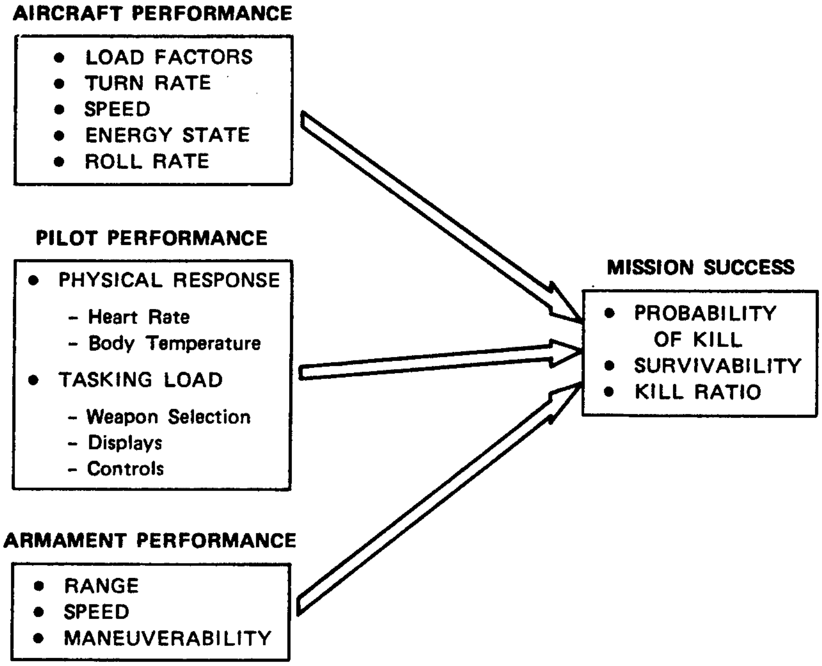


FIGURE 20 SYSTEM PERFORMANCE ASSESSMENT

Specifically, the instrumentation systems described could significantly add to the understanding of the following questions:

o How can real combat system performance be assessed on an operational training range in terms of kill probability, survivability, and kill ratio?

o What is the weak link in the airborne weapon system--the aircraft, the pilot, or the armament?

o If pilot workload is the limiting factor, which component is most influenced--the mental workload or the physical workload?

The TACTS/ACMI and the APMS instrumentation systems together can answer these questions.

REFERENCES

1. S. A. Rositano, Objective Measurement of Human Tolerance to +Gz Acceleration Stress, NASA Technical Memo 81166 (February 1980).

2. J. Nyboer, Electrical Impedence Plethysmography Second Edition (Charles C. Thomas, Publisher, Springfield, Illinois; 1970).

3. L. D. Montgomery, "Comparison of an Impedance Device to a Displacement Plethysmography for study of Finger Blood Flow," Aviation, Space, and Environmental Medicine, Vol. 47, No. 1, pp. 33-38 (1976).

4. J. C. Dennifor et al., "Measurement of Cardiac Output by Electrical Impedance at Rest and During Exercise," Journal of Applied Physiology, Vol. 40, No. 1 (January 1976).

5. C. Z. Naggar et al., "Accuracy of the Stroke Index as Determined by the Transthoracic Electrical Impedance Method," Anesthesiology, Vol. 42, No. 2 (February 1975).

6. J. McHenry and R. L. York, "A New Test Arena for Aircraft Research and Development," presented at AIAA Aircraft Systems Meeting 4-6 August 1980, Anaheim, California.

7. R. L. York, D. K. Rust, C. W. Terndrup, I. M. Wadsworth, "Analysis and Enhancement of TACTS Data for Application to the Study of Airframe Structural Fatigue," SRI Memorandum 2950-81-RM-179 (December 1981).

HUMAN FACTORS IN NIGHT ATTACK - TEST METHODOLOGY

BY

L/C Paul M. Lang, HQ AFTEC
Dr. J. D. Boren, BDM-Albuquerque
Mr. Fred Ragland, BDM-Albuquerque

TITLE This paper presents an interim report on a study designed to assess the overall operational effectiveness of a night attack weapon system when primary areas of concern include aircrew workload and stress.

PURPOSE Night attack conducted at low-altitude is a primary thrust of many developing weapon systems (the B-1, F-16, MSIP, F-16E, F-15E, the H-X, to name a few). In many of these systems concern exists as to whether aircrew tasking is being pushed into an area when human restrictions will impose serious limits upon operational effectiveness. Division of attention (mental and physical), reaction time, and stress are all becoming more critical in this mission area. There is a serious DOD-wide need to eliminate the risks and uncertainty in these programs that arise from these human factors concerns. The answer has high value if found early in these programs and high cost if delayed.

APPROACH The study approach consisted of three broad steps. The first was to lay out the critical issues that arise in this mission area for any manned system. Our underlying concept of the aircrew is that of an individual with flexible but limited capacity for sensory-motor skills, mental calculations/decision-making, and ability to function under stress. These human capacities, different for each individual, interact with specific tasks generated by the characteristics of the weapon system and mission requirements. Any test for operational effectiveness must address these three factors:

INDIVIDUAL VARIATIONS

WEAPON SYSTEM CHARACTERISTICS

MISSION TASKING/ENVIRONMENT

Our second phase was the establishment of an overall methodology and data requirements based upon these three functions.

The third step was a survey of available methods/facilities/equipment for the execution of such a test.

CRITICAL DOD Directive 5000.3 defines Operational Test and Evaluation as:

> "That test and evaluation conducted to estimate a system operational effectiveness and suitability, as well as the need for any modification. It is accomplished by operational and support personnel of the type and qualifications expected to use the system when deployed and is conducted in as realistic an operational environment as possible."

<u>Operational Effectiveness</u> is defined by AFR 80-4 as:

> "How well the system performs its intended mission in its intended environment."

These meanings translate into total man/machine system capability to penetrate enemy defenses and countermeasures in such a manner that an appropriate number of targets can be killed at acceptable levels of sortie effort and attrition losses. Therefore, critical components of the test include:

REALISTIC TERRAIN/TARGETS

DAY/NIGHT OPERATIONS

ADVERSE WEATHER

INTENSE DEFENSIVE ARRAY

ENVISIONED COUNTERMEASURES

appropriate for the operational life of the system.

STRETCHING THE MAN IN NIGHT ATTACH

The cumulative burden placed on the human operator can be envisioned as the three dimensional integral volume under a surface determined by the level of performance in each task over the number of tasks or "workload" and the physical and psychological stress effective during the tasks. For each mission there is a "demand surface," and for each individual aircrew/mission combination, there is a "capability surface." The question, therefore, to be answered is whether capability is above or below mission demands and what state the human operator is in while performing the mission.

The night attack mission pushes the aircrew beyond previous operational tasking in several respects:

1. It requires precise split-second nap of the earth flying.
2. It requires the monitoring of multiple sensors (radar, RHAW gear, FLIR, etc.).
3. It removes normal cues previously utilized for low-altitude flying.
4. It projects the system deep into intense defenses and new countermeasures.

SAMPLING STRATEGY

The operational effectiveness test must subject a representative cross-section of intended human operators to realistic mission tasking and environment. Various experimental controls must be employed to insure that the test characterizes the projected operational effectiveness of a "typical" unit in the mission tasking. Experience, training, and response to stress must be used in selecting test aircrews. Sufficient subjects must be included to represent individual variations.

SYSTEM PERFORMANCE ASSESSMENT

The interactions of the man/machine/mission/environment are complex. Yet, they must be faced in their synergistic combination since that interaction is the central issue of concern.

In addition to variations in the human operator's capability, there are natural operating factors which interact to drive operational effectiveness. Key among these are the trade offs between terrain clearance and speed utilized to minimize attrition. The low-altitude contest between penetrator and defenses is one of the major evaluations of modern tactical warfare.

High-speed, very low-altitude flight is possible over smooth terrain, but defenses can also operate very low under these circumstances. Encountering rough terrain will force either a climb or reduction in speed. All these variations are impacted by visibility conditions and the availability of peripheral cues (ambient light).

For system's effectiveness, these factors must all be considered while still summarizing to an overall warfighting capability and combat loss rate.

EASURES F FFECTIVENESS

In judging overall weapon system effectiveness when the primary concern is human limitation, we not only assess performance in the ability to accurately find and kill appropriate numbers and types of targets, or survivability rates from various causes, but we must also be aware of the internal state of the human operator (for both reliability and combat sustainability reasons).

Three strategies are available.

I - The first is to assess performance by observing differences of operators. However, we have no clue how close to human limitations we are operating or what type of deterioration might occur if these limits are exceeded.

II - The second strategy is to observe performance while measuring the reserve capacity in cognitive, sensory-motor, decision tasks, as well as physical and psychological stress. This strategy leads to some expectation in combat sustainability.

III- The third strategy is to have some criteria in reserve capacity to insure sustainability and then see what combat performance is still attainable.

Both II and III are desirable strategies for OT&E.

HUMAN FACTORS DATA

In light of the previous discussion, the data requirements are driven by the need to characterize operational measures of effectiveness as functions of the component capabilities of the individual. Knowing these relationships for various test conditions along a scale of workload/stress combinations provides a true insight into the state of the individual as he meets these varied demands. These trends also make possible the last critical step, that is the characterization of potential combat effectiveness based upon test conditions that can never reach the full severity of the next war. In other words, we need to extrapolate to WWIII as best we can.

TEST CONDITIONS

If we are to accurately characterize the man/machine system, every key system function must be duplicated in the test precisely as it interacts with the man and the mission environment. Further, and of equal importance, the test must provide realistic mission tasking in terms of terrain, weather, ambient light, distance to target, etc., as well as the critical interaction with enemy forces. Finally, the proper instrumentation, test techniques and recording equipment must be set up to capture both performance and human state.

EXPERIMENTAL DESIGN AND CONTROL

Test events need to be controlled and repeated enough times such that the functional relationship between performance and human capabilities can be confidently reported. Only when adequate data is gathered regarding each subject and the potential user population is adequately represented by enough subjects, can we report performance expected from personnel "of the type and qualifications expected to use the system when deployed."

EVALUATION OF POTENTIAL METHODS

Having thus established the requirements and characteristics of the desired test we now turn to an evaluation of different types of test execution.

Flight test includes airborne simulation, specifically designed test vehicles, prototypes, and pre-production hardware.

Flight simulator is a ground based simulation centered around the crew position with appropriate displays, controls, video, computer support and scenario/threat data bases.

Task Analysis (TA) is basically an action path chart on paper of the tasks/decisions required to perform the mission.

Computer is essentially the interaction of a "computer man" with a TA input.

Questionnaires include various subjective methods which require the human operators to "self-evaluate" his experiences and mental state.

QUANTIFYING PERFORMANCE MOE'S

The essential issue then for operational effectiveness testing is the ability of these five classes of tests to provide quantitative data for the operational performance MOE's either directly, by inferrence, or by providing supporting information. Additional data on the simultaneous state of the operator is also desirable. On the chart an "X" indicates inability to credibly address the MOE with that method. Other comments within the chart are self-explanatory.

Our evaluation concluded that only flight test and flight simulators can credibly provide the performance measures needed to make an overall system evaluation. This conclusion is reinforced when considering the major issue to be the cumulative interaction of mission tasking on the man/machine system. Many synergetic effects are likely in this complex scenario which can only be addressed by subjecting potential operators to the multiple tasks required. When the issues at hand include division of attention, situation awareness, and attention allocation strategy under a complex mission tasking/stress, it is recommended to test to that total load and stress as closely as practical.

The fundamental issue remains that short of full automation, each of these MOE's is determined by the actions/decisions of the operator.

Essential elements of the test include:

Visual crosscheck outside and inside the cockpit.
Auditing perception/feedback.
Overall situation awareness/staying ahead of events.
Mental calculating and decision time.
Piloting skill (quickness + firmness of control).
Manual dexterity/switchology.
Cockpit control/display integration.
Aircrew psychological makeup.

It is obvious from this tasking that training and experience of the aircrew are key factors in this evaluation.

Therefore, two other issues must be integrated into the test design. First, a representative cross-section of potential operators must be included in the evaluation. Second, a history of performance versus experience within the test must be kept as an indication of training requirements when the system is deployed.

Thus, the quantifying of operational effectiveness, and the interaction of training together lead to in-flight testing and ground based simulation as the two primary methods to be used.

COMPARISON OF TEST METHODS

A comparison of flight test versus ground based simulation is a study in technological and cost limitations rather than inherent restraints. In applying the data requirements against these two methods, we find the following capabilities.

WEAPON SYSTEM FIDELITY. The ability to reproduce avionics signal processing is really only dependent upon generating those signals or a reasonable facsimile from the simulation. Thus, a terrain following radar cue to the pilot can be the result of actual radar returns in flight or the terrain data base of a digital simulation as unmasked from present position coordinates. The cost may be lower for a ground based simulator since only the logic and display are necessary and electronic equipment does not require in-flight rating. One key area of technological concern for the night mission is the ability to generate a realistic infrared video display that truly interacts with cloud/humidity/aerosol content as selected from the simulator sonsole.

SCENARIO FIDELITY. The ability to duplicate World War III or a NATO conventional conflict does not currently exist for either method. For in-flight testing true nap of the earth flying is limited by safety constraints. Defense simulations or captured equipment is available for flight test but not in sufficient quantity/quality to truly simulate a Warsaw Pact defensive array either at the FEBA or at a high value interdiction/strike target.

For a ground based simulation, the terrain can be generated either by scale model board or computer generated imagery to duplicate actual European geography. Model boards are severely limited in flight duration, in that flight at fighter/bomber speeds quickly consumes the available distance in a matter of a few minutes. Computer generated imagery (CGI) has progressed from a very rough approximation of terrain with few enhancing aspects (trees, roads, etc.) to textured imagery with a sizeable menu of enhancing objects on call CGI has no real distance limitation other than the availability of topographic data. Defensive arrays are also limited only by computer memory and processing power. As defensive sites (SAM or AAA) are unmasked, their characteristics can be used to activate cockpit radar homing and warning (RHAW) gear indications, and interactive engagements can be played to assess the probability of kill.

Finally, areas basically unsafe for flight test can be investigated in simulators. These include serious overload, fatigue, and potentially life threatening ground clearances.

Flight does produce unique physical sensations that cannot be reproduced on the ground. Attempts to generate motion cues in simulators physically with hydraulic systems have generally been very costly and of limited use due to limited duration and negative cues. Thus, other motion cues due to vision, inflation of "g" suits, etc., have come into use.

Finally, the question of whether the subject becomes wholely convinced and excited during testing such that combat levels of emotional stress are experienced, has caused doubt in the results of operational testing. Emotional stress can induce heightened or suppressed performance. Further, these effects occur under different conditions for different individuals, or even the same individuals at different times. Various stress indicators such as pulse, blood pressure, respiration rate, voice, etc., are known. It is reasonable, therefore, to gather readings on these indications both in test and in combat. Over an extended period, a functional pattern could be established between these indicators and performance. It is felt that performance enhancements or degradations at various levels of stress as a pattern are far more important data than any single reading obtained by trying to artifically duplicate fear or anxiety.

MULTIPLE POINT PROTECTION

By measuring performance MOE's under various levels of stress indicators for each human operator, multivariate analysis can be used to predict combat performance. Due to the complexity of the interactions and the influence of individual differences, a significant test matrix of individuals, test conditions, and controlled replications must be executed. Only then can the emerging factors influencing operational effectiveness be defined. Multiple analytical techniques can be used against these data including:

Simple trend analysis
Constant 2nd/3rd derivative assumptions and stepwise integration
Quadratic function definition by multiple points
Curve fitting
Correlation/Regression Analysis
Multivariate Analysis

These techniques call for the use of both ground based simulation and flight test.

ADVANTAGES OF COMBINED METHODS

It is therefore the conclusion of this study that both ground based simulation and flight test are required for the operational effectiveness evaluation of a complete weapon system.

Currently, simulators have significant advantages in safety, cost, experimental control, and scenario intensity. Flight test has the advantage in ultimate realism, motion, and weapons employment.

By using both in an organized, sequential program the strongest features of each method can be applied. These mutually complimentary data gathering strengths not only complete the loopholes in a single method, but generate an interlocking pattern which can be used to project beyond test conditions toward combat expectations.

Past studies have shown a sortie cost comparison of simulator versus flight of one to six. Thus, if the data from each were equally valued, a typical test would tend to have more simulator sorties by the inverse ratio. Time and cost pressures could push the ratio higher.

In any program, the use of both methods would be more cost effective than reliance on either method by itself, by leading to more complete data, and enhanced use of resources.

PURPOSE

- DoD WIDE REQUIREMENT FOR NIGHT/ALL WEATHER ATTACK
- CAN IT BE DONE?
- IS WORKLOAD EXCESSIVE?
- AFTEC NEED FOR EARLY DETERMINATION
- HIGH PAYOFF

APPROACH

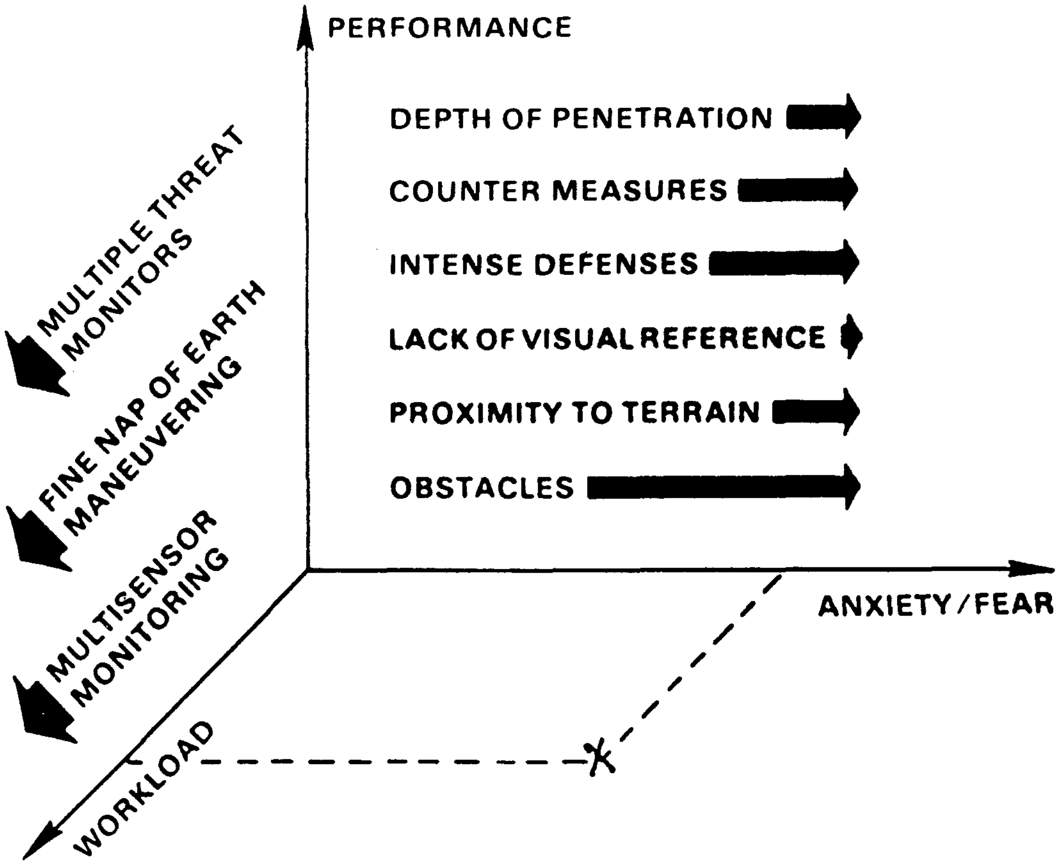
- DEVELOP HUMAN FACTORS TEST REQUIREMENTS
  - CRITICAL QUESTIONS
  - SURVEYED HUMAN CAPABILTIES
  - MATCH PILOT CAPABILITIES WITH SPECIFIC MISSION TASKS
  - ASSESS MAN/MACHINE INTERACTION
- ESTABLISH TEST METHODOLOGY
  - HUMAN FACTORS DATA REQUIREMENTS
  - SYSTEMS, SCENARIO CHARACTERISTICS
  - EXPERIMENTAL DESIGN AND CONTROL
- SURVEY OF AVAILABLE METHODS

CRITICAL QUESTIONS

1. CAN THE AIRCREW, IN THE ENVISIONED ENROUTE THREAT ENVIRONMENT, PERFORM LOW ALTITUDE NAVIGATION TO A TARGET DURING DAY/NIGHT, ADVERSE WEATHER CONDITIONS?

2. CAN THE AIRCREW, IN THE ENVISIONED TERMINAL THREAT ENVIRONMENT, PERFORM EFFECTIVE ATTACKS ON APPROPRIATE TARGETS DURING DAY/NIGHT, ADVERSE WEATHER CONDITIONS?

STRETCHING THE MAN IN NIGHT ATTACK



PERFORMANCE

DEPTH OF PENETRATION

COUNTER MEASURES

INTENSE DEFENSES

LACK OF VISUAL REFERENCE

PROXIMITY TO TERRAIN

OBSTACLES

ANXIETY/FEAR

WORKLOAD

MULTIPLE THREAT MONITORS

FINE NAP OF EARTH MANEUVERING

MULTISENSOR MONITORING

SAMPLING STRATEGY

SUBJECTS MUST BE:

REPRESENTATIVE IN
- EXPERIENCE
- SPECIFIC TRAINING

QUANTITY
- SUFFICIENT TO COVER — DIVERSITY
  — ACCURACY
  — CONFIDENCE LEVEL

SYSTEM PERFORMANCE ASSESSMENT

- COMPLEX INTERACTIONS DETERMINE OUTPUTS

INPUTS	FACTORS WHICH INFLUENCE	OUTPUTS
MISSION TASKING	ALTITUDE	
SYSTEM PERFORMANCE	AIRSPEED	PROBABILITY OF MISSION SUCCESS
SYSTEM INTEGRATION	TERRAIN	
INDIVIDUAL PILOT ABILITY — SKILL — WORKLOAD — ANXIETY — STRESS	DEFENSES AMBIENT LIGHT WEATHER	PROBABILITY OF SURVIVAL

MEASURES OF EFFECTIVENESS

A. MISSION PERFORMANCE
  - NAVIGATION ACCURACY
  - TARGET ACQUISITION/KILL

B. SURVIVABILITY
  - GROUND IMPACT
  - SHOOT DOWN

C. ADDITIONAL PILOT CAPABILITY TO PERFORM
  - NORMAL
  - EMERGENCY

HUMAN FACTORS DATA

A. MEASURE THE OPERATIONAL MOE's UNDER:
  - VARIOUS DEGREES OF ''BUSYNESS''/WORKLOAD
  - DIFFERENT TYPES AND DEGREES OF PHYSICAL STRESS/SENSATIONS
  - DIFFERENT TYPES AND DEGREES OF PSYCHOLOGICAL STRESS/ANXIETY

B. ESTABLISH CORRELATION PATTERN BETWEEN HUMAN FACTORS AND OPERATIONAL PERFORMANCE

C. PROJECT TEST PERFORMANCE TO ANTICIPATED SUCCESS/SURVIVAL

TEST CONDITIONS

A. HIGH FIDELITY WEAPON SYSTEM
  - DISPLAY ENVIRONMENT INTERACTIONS
  - NAVIGATION COMPUTER
  - TERRAIN FOLLOWING/AVOIDANCE
  - FIRE CONTROL
  - RHAW/ECM
  - WEAPONS EFFECTS
  - COCKPIT LAYOUT/INTEGRATION
  - OTHER

B. HIGH FIDELITY OPERATIONAL SCENARIO
  - TERRAIN
  - WEATHER AND TIME-OF-DAY EFFECTS
  - DEFENSES (SAM, AAA, INTERCEPTOR)
  - TARGET/DEFENSE DEPLOYMENT

C. HUMAN FACTORS DATA RETRIEVAL

D. PERFORMANCE DATA RETRIEVAL

EXPERIMENTAL DESIGN AND CONTROL

A. PILOT WORKLOAD COMPONENTS

B. TASK DIFFICULTY CONDITIONS

C. NUMBER OF REPLICATIONS

D. SELECTION OF SUBJECTS

EVALUATION OF POTENTIAL METHODS

	FLT TESTS	FLT SIM	TA	COMPUTER	QUESTIONNAIRE
TEST REQUIREMENTS WORKLOAD ANXIETY PHYSICAL STRESS PERFORMANCE RELATE TO COMBAT					
TEST CHARACTERISTICS WPN SYSTEM FIDELITY SCENARIO FIDELITY HF SUPPORT					
EXPERIMENTAL CONTROL LOAD COMPONENTS TEST CONDITIONS NUMBER OF REPLICATIONS SELECTION OF SUBJECTS					

QUANTIFYING PERFORMANCE MOEs

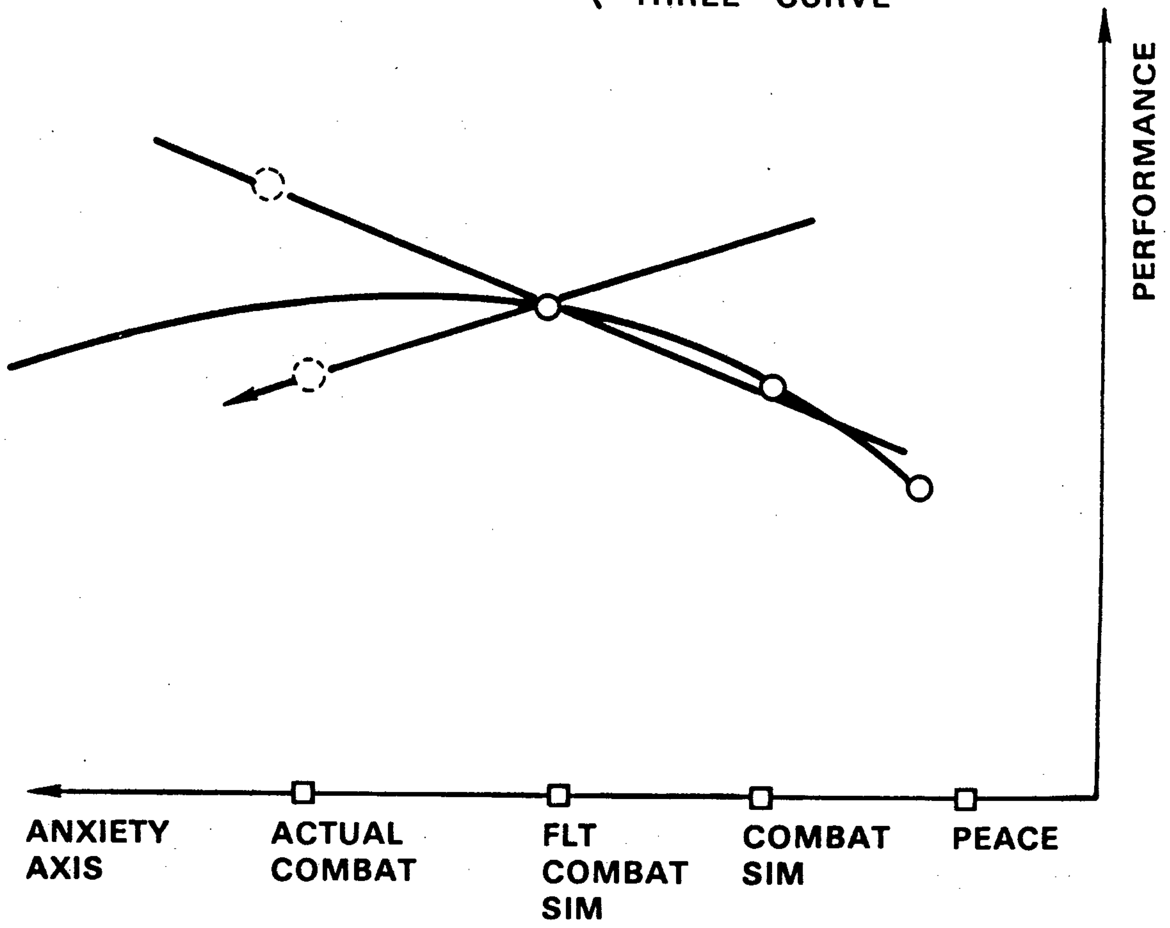
	FLT TEST	FLT SIM	TASK ANALYSIS	COMPUTER	QUESTIONNAIRE
NAVIGATION ACCURACY/TGT ACQUISITION	DIRECT TSPI	DIRECT TSP	INFERRED TIME DEVOTED	INFERRED TIME DEVOTED	COULD PROVIDE INPUT
TGT KILL	RANGE EVENTS	PREDICTED IMPACT	INFERRED TIME DEVOTED	INFERRED TIME DEVOTED	COULD PROVIDE INPUT
GROUND IMPACTS	UNSAFE	DIRECT OBSERVATION			COULD PROVIDE INPUT
FLY UPS	DIRECT TSPI	DIRECT TSPI			COULD PROVIDE INPUT
SURVIVABILITY	INFERRED FROM MODEL. INTERACTIVE IN LARGE EXERCISES	INFERRED FROM MODEL. INTERACTIVE INDICATIONS IN SIM.			COULD PROVIDE INPUT
ERRORS	INFERRED OR RECORDED	DIRECT RECORDING	INFERRED	INFERRED	COULD PROVIDE INPUT

COMPARISON OF TEST METHODS

	FLT	SIM
WPN SYSTEM FIDELITY	BOTH CAN CONTAIN APPROPRIATE COCKPIT/AVIONICS. SIM COULD BE MODIFIED AND RUN AT LOWER COST.	
SCENARIO FIDELITY		
TERRAIN	CAN GET CLOSE BUT PEACETIME FLYING RESTRICTIONS APPLY.	CAN BE IDENTICAL TO EUROPE DIGITAL DMA DATA
DEFENSES	CAN ONLY GET PARTIAL REALISM IN LARGE EXERCISES	CAN DUPLICATE DENSE ARRAYS AND DEAL INTERACTIVELY
SURGE, ETC.	CAN DO BUT NOT SAFELY	
H. F. DATA WORKLOAD	BOTH CAN MAKE THE MAN AS BUSY AS NECESSARY (BUT SAFETY MAY BE AFFECTED IN FLIGHT)	
PHYSICAL STRESS	EXCELLENT	ONLY CUES
ANXIETY FEAR	JUST BEING IN A COCKPIT OR UP IN THE AIR DOES NOT CREATE ANXIETY OF COMBAT	
	FEAR OF HITTING GROUND PEACETIME SAFETY/ ACCURACY OF DATA	FEAR/PRIDE IN THE SCENARIO. NO SAFETY LIMITS/ACTUAL HUMAN CAPABILITY

SINGLE PERFORMANCE POINT CANNOT TELL STORY BUT TWO OR THREE SET A PATTERN TO PROJECT

MATHEMATICALLY 〈 TWO = SLOPE
THREE = CURVE



BLOOD, URINE, VOICE, ETC., TELL US HOW FAR ALONG ANXIETY AXIS WE ARE

ADVANTAGES OF COMBINED METHODS

- "STRONGEST" FEATURES OF INDIVIDUAL METHODS.
- ALLOWS "HOLES" IN THE DATA AVAILABLE FROM A SINGLE METHOD TO BE FILLED
- ABILITY TO CROSS CORRELATE OR ESTABLISH TRENDS BETWEEN METHODS USING MEASUREABLE DIFFERENCES IN KEY FACTORS
- ALLOWS INCREMENTAL COST AND RISK
- ENHANCED PLANNING/REPORTING

THEORETICAL BASIS FOR WORKLOAD ASSESSMENT RESEARCH AT NASA-AMES RESEARCH CENTER

Sandra G. Hart
NASA-Ames Research Center
Moffett Field, CA

SUMMARY

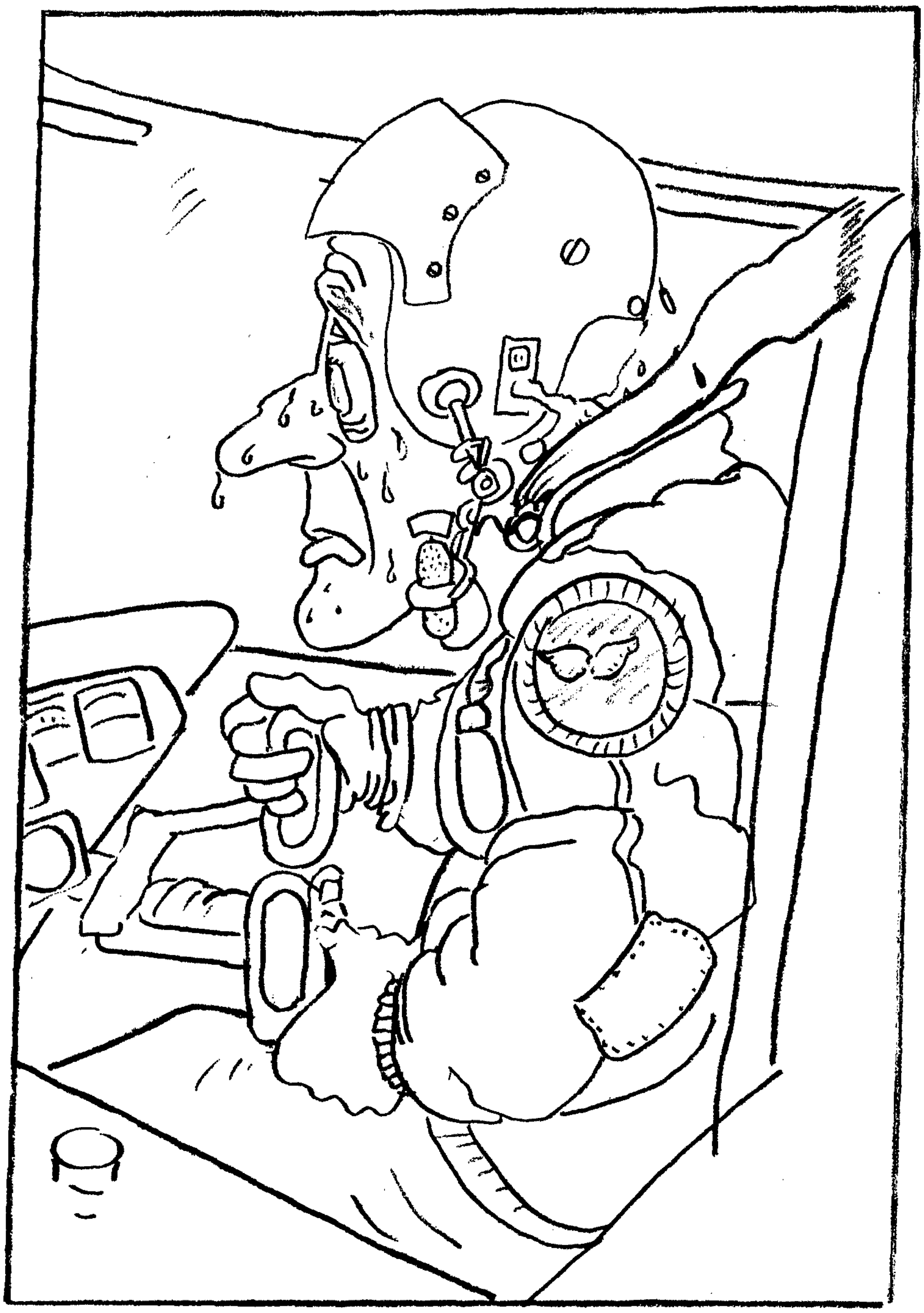
Workload may be thought of as a collection of experiences, requirements, feelings, demands, and circumstances that are referred to in summary form by the term "workload". When one person says that he really worked hard, he may mean that he is physically tired, while another person may provide a rating of equivalent magnitude because he was required to do more than expected, even though his actual output and effort did not increase. There are many factors associated with the term workload as it is usually applied that each exist independently and can be analyzed as such most profitably. Task demands are just that - - task demands. No additional meaning or value can be associated with renaming this factor "workload". Physical effort and emotional stress are also independent, unique entities that can each be measured by specific and unique assessment techniques, but again neither is synonomous with "workload" per se. Performance is also an independent, important entity, but again it is not "workload". Measures of performance are most relevant to determining how successful an individual was in meeting task demands but do not reflect how hard he worked, what his expectations were, his stress level, the time pressure felt, and so on.

The one factor that does reflect the effect of all of these factors on each individual is the subjective experience of workload. If an individual feels loaded, he or she is. This may be the only factor in the constellation of elements variously called "workload" that is purely "workload" and nothing else. This subjective experience is obviously derived from the other factors - - task demands, success in meeting demands, effort, and so on - - but it is the product of a weighting process that may be unique to each individual. The weights or importance that each individual places on the various elements that may affect his experience of workload may differ from person to person, although they should be fairly consistent within an individual. For example one individual may consider his workload to be defined by the demands of a task whereas another person may only rate workload as high if he is physically tired, and yet another might primarily base his estimate on his ability to meet the demands of a task, rather than the amount of the demands or the effort exerted.

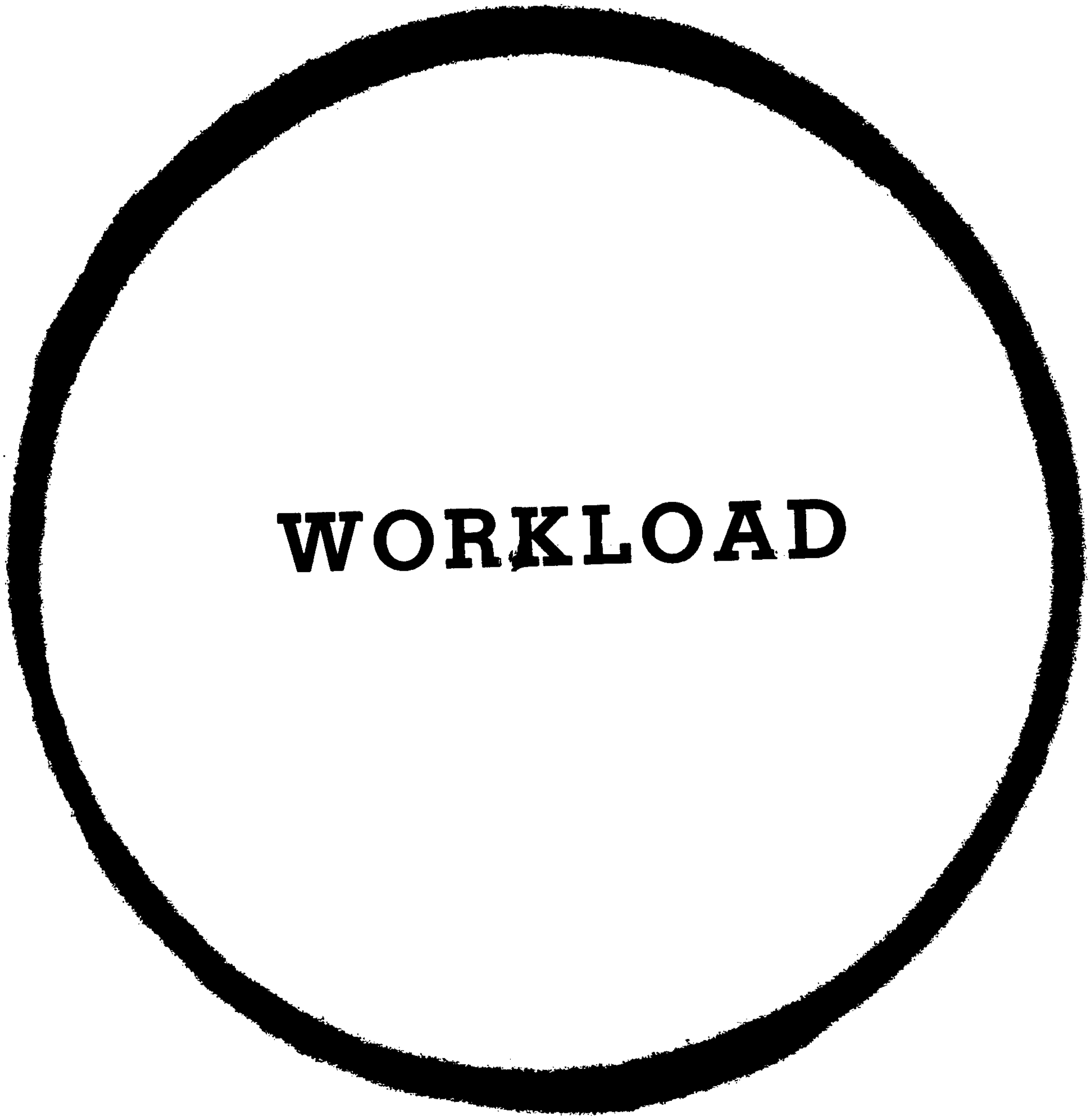
By determining what factors enter into this weighting process and how they are combined, it may be possible to develop methods to assess this subjective factor - - the one element that may be uniquely "workload" - - to use in the interpretation of subjective ratings, variation in performance, and physiological recordings. The assumption is that if a person feels loaded - - he is - - and that this will not only affect his or her subjective evaluations of workload but also physiological measures of stress, arousal, fatigue, etc. and the individual's ability to perform the primary task as well as additional tasks effectively.

"WORKLOAD"

WORKLOAD IS A SUBJECTIVE EXPERIENCE CAUSED BY EXTERNAL AND INTERNAL FACTORS SUCH AS MOTIVATION, ABILITY, EXPECTATIONS, TRAINING, TIMING, STRESS, FATIGUE, AND CIRCUMSTANCES IN ADDITION TO THE NUMBER, TYPE, AND DIFFICULTY OF TASKS PERFORMED, EFFORT EXPENDED, AND SUCCESS IN MEETING REQUIREMENTS.

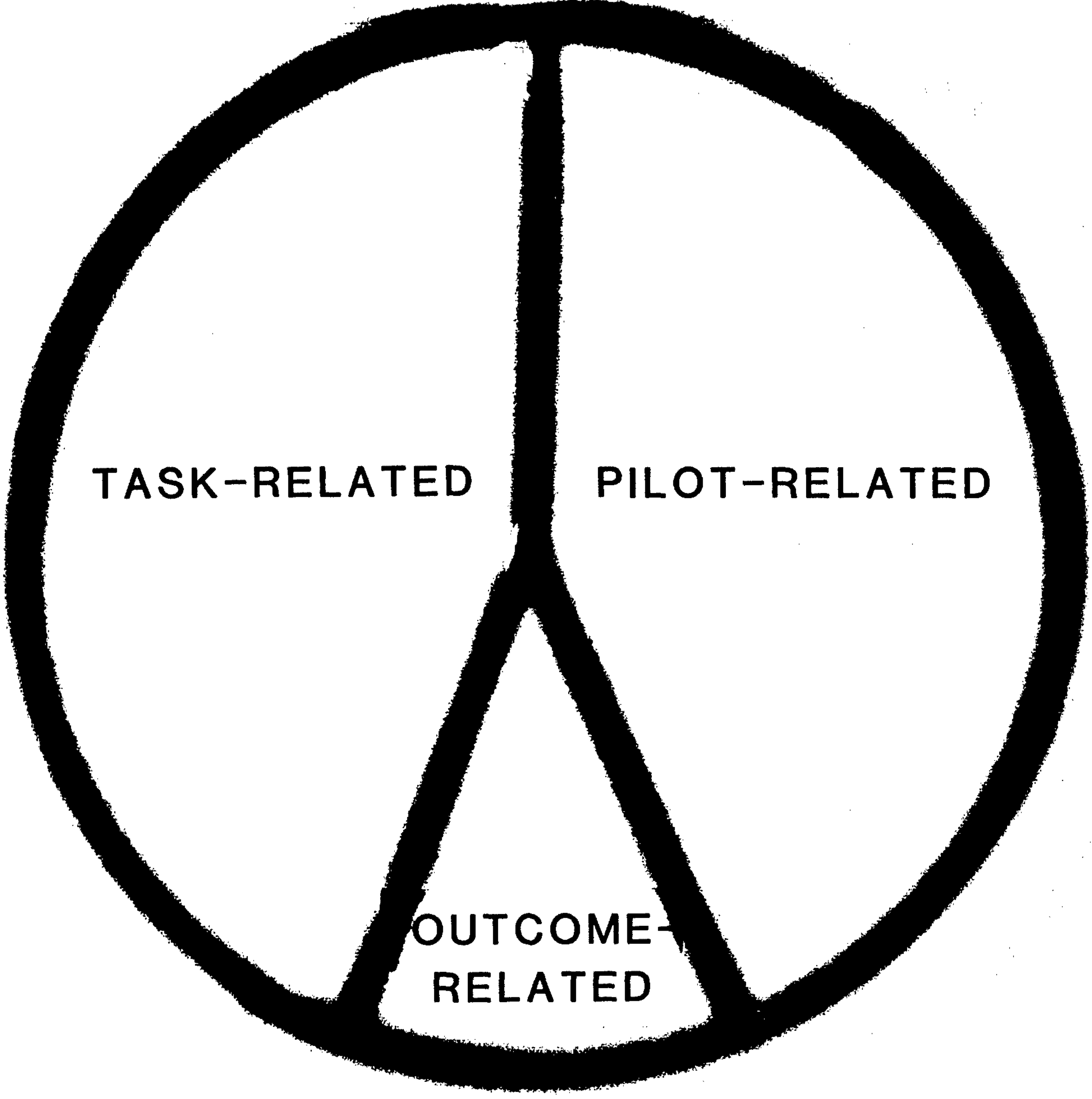


A SUBJECTIVE EXPERIENCE OF WORKLOAD



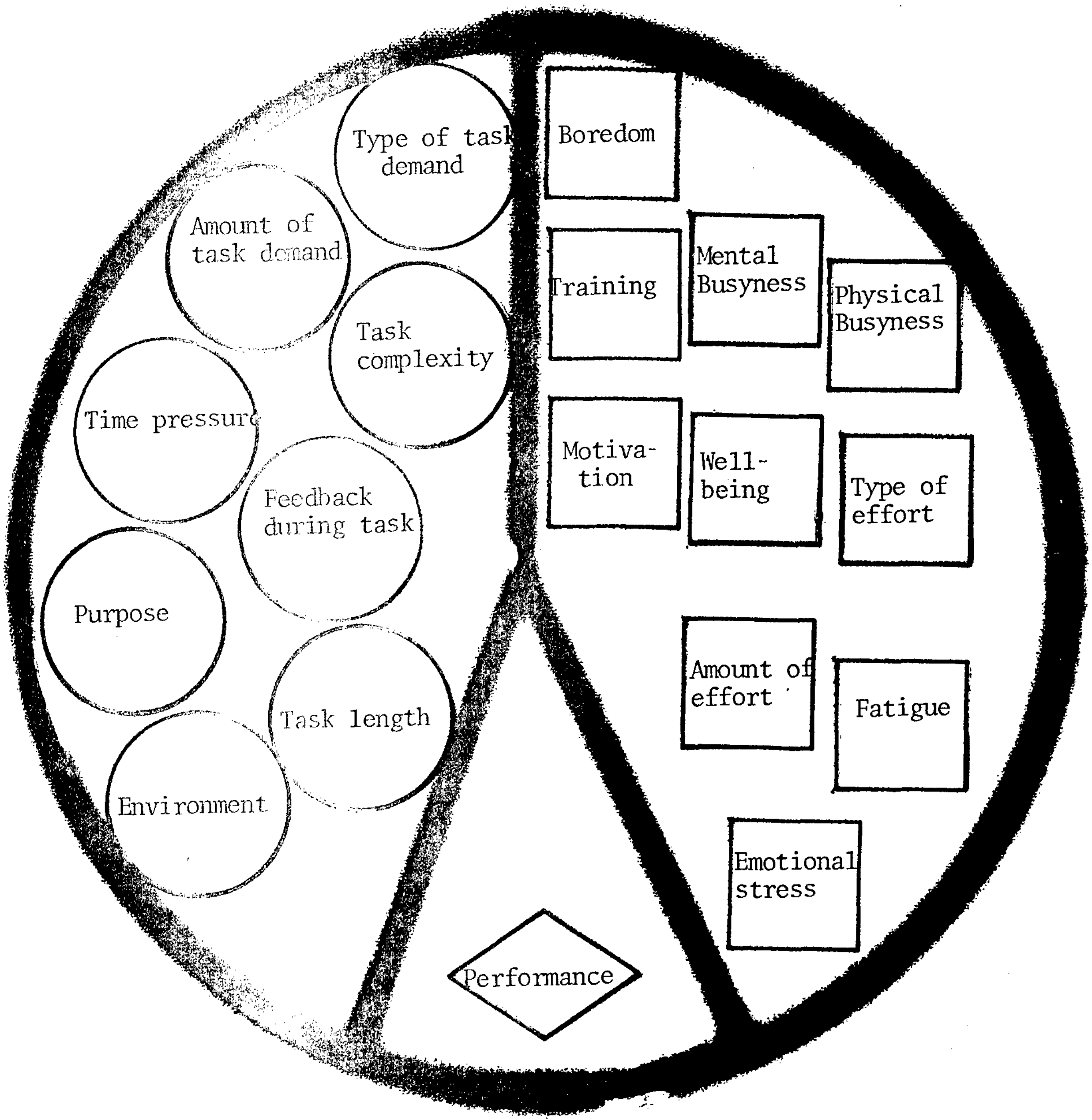
THE AREA OF THE ABOVE FIGURE REPRESENTS A SUBJECTIVE EXPERIENCE OF WORKLOAD AS MEASURED BY A RATING, OBJECTIVE MEASURE OR PHYSIOLOGICAL INDICE.

THERE ARE SEVERAL DIFFERENT DIMENSIONS THAT COMBINE TO CREATE THE EXPERIENCE OF WORKLOAD



THE AREA OF THE CIRCLE REPRESENTS THE MAGNITUDE OF WORKLOAD EXPERIENCED
THE DIFFERENT AREAS AND COLORS REPRESENT THE DIMENSIONS OF THAT EXPERIENCE

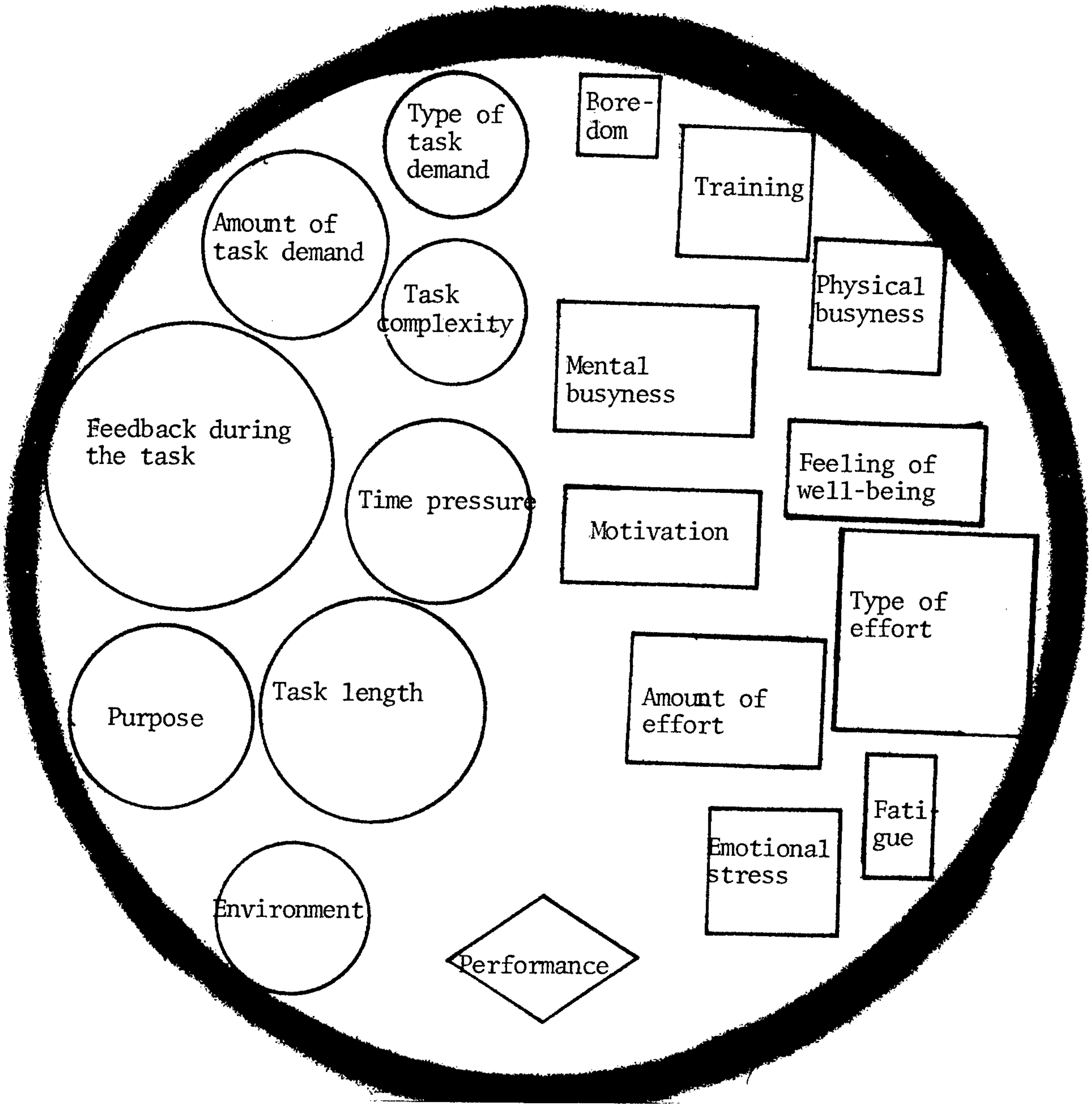
COMPONENTS OF THE DIMENSIONS THAT CREATE THE EXPERIENCE OF WORKLOAD



THE AREA OF THE CIRCLE REPRESENTS THE MAGNITUDE OF WORKLOAD EXPERIENCED

THE SHAPE AND COLOR OF THE COMPONENTS REPRESENTS THEIR SOURCE (TASK/OPERATOR/OUTCOME)

THE MAGNITUDE AND NATURE OF THE COMPONENTS MAY VARY



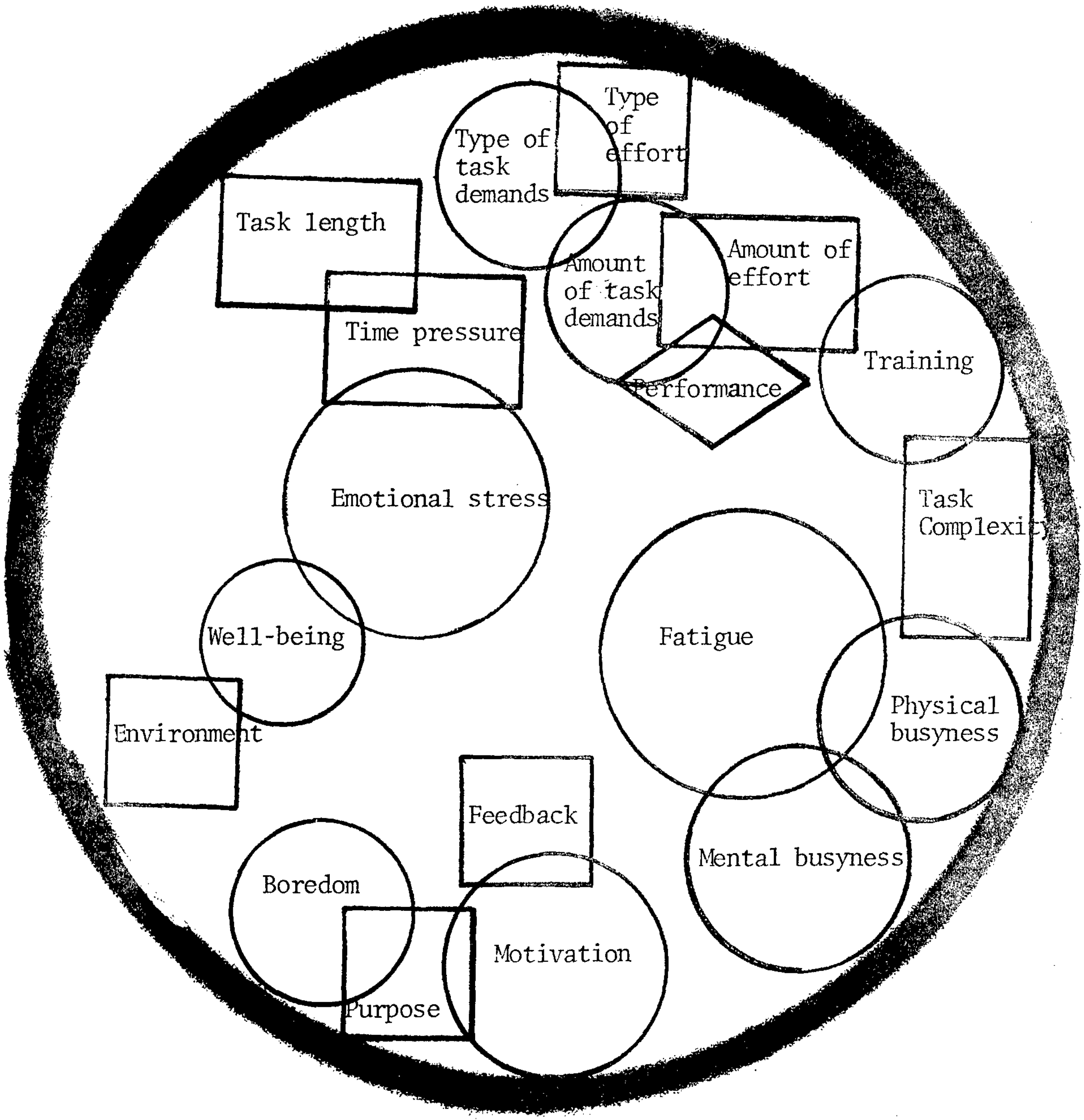
THE AREA OF THE CIRCLE REPRESENTS THE MAGNITUDE OF WORKLOAD EXPERIENCED

THE SIZE OF EACH COMPONENT INDICATES HOW MUCH OF IT WAS PRESENT AND RELEVANT IN A GIVEN EXPERIENCE

THE SHAPE AND COLOR OF EACH COMPONENT REPRESENTS ITS SOURCE (TASK- OPERATOR- OR OUTCOME-RELATED)

THE COMPONENTS MAY OR MAY NOT COVARY

(THESE INTERRELATIONSHIPS MAY BE CAUSAL OR COINCIDENTAL, POSITIVE OR NEGATIVE)



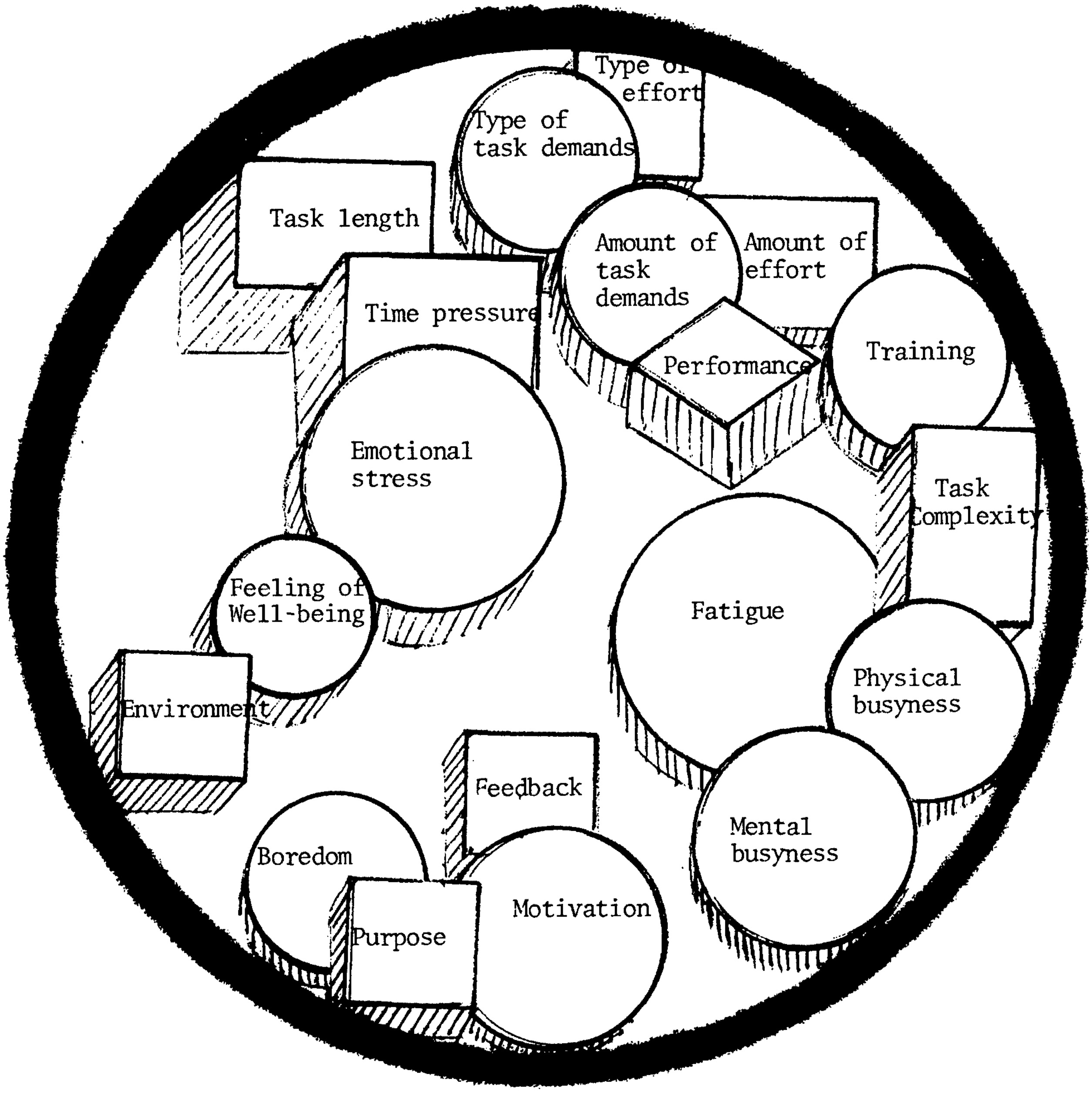
THE SHAPE/COLOR OF A COMPONENT REPRESENTS ITS SOURCE(TASK- OPERATOR- OR OUTCOME-RELATED)

THE SIZE OF THE COMPONENT REPRESENTS THE "AMOUNT" OF THAT COMPONENT PRESENT

THE INTERSECTION BETWEEN COMPONENTS REPRESENTS COVARIATION

THE AREA OF THE CIRCLE REPRESENTS THE MAGNITUDE OF THE WORKLOAD EXPERIENCED

THE IMPORTANCE GIVEN TO EACH COMPONENT IN ESTIMATING OR REACTING TO WORKLOAD ALSO VARIES



THE SHAPE/COLOR OF COMPONENTS REPRESENTS SOURCE (TASK/OPERATOR/OUTCOME)

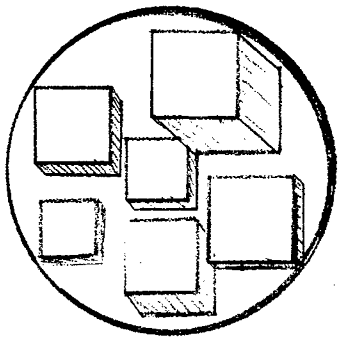
THE SIZE REPRESENTS THE "AMOUNT" OF THAT COMPONENT PRESENT

THE INTERSECTION BETWEEN COMPONENTS REPRESENTS COVARIATION

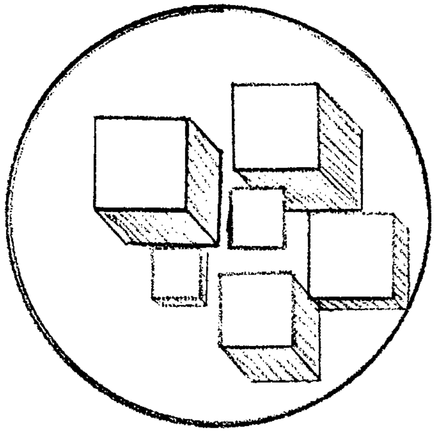
THE HEIGHT OF COMPONENTS REPRESENTS THE IMPORTANCE OR WEIGHT THEY ARE GIVEN BY ONE INDIVIDUAL

THE AREA OF THE CIRCLE REPRESENTS THE MAGNITUDE OF THE WORKLOAD EXPERIENCED

WORKLOAD IS RATED AS LOW

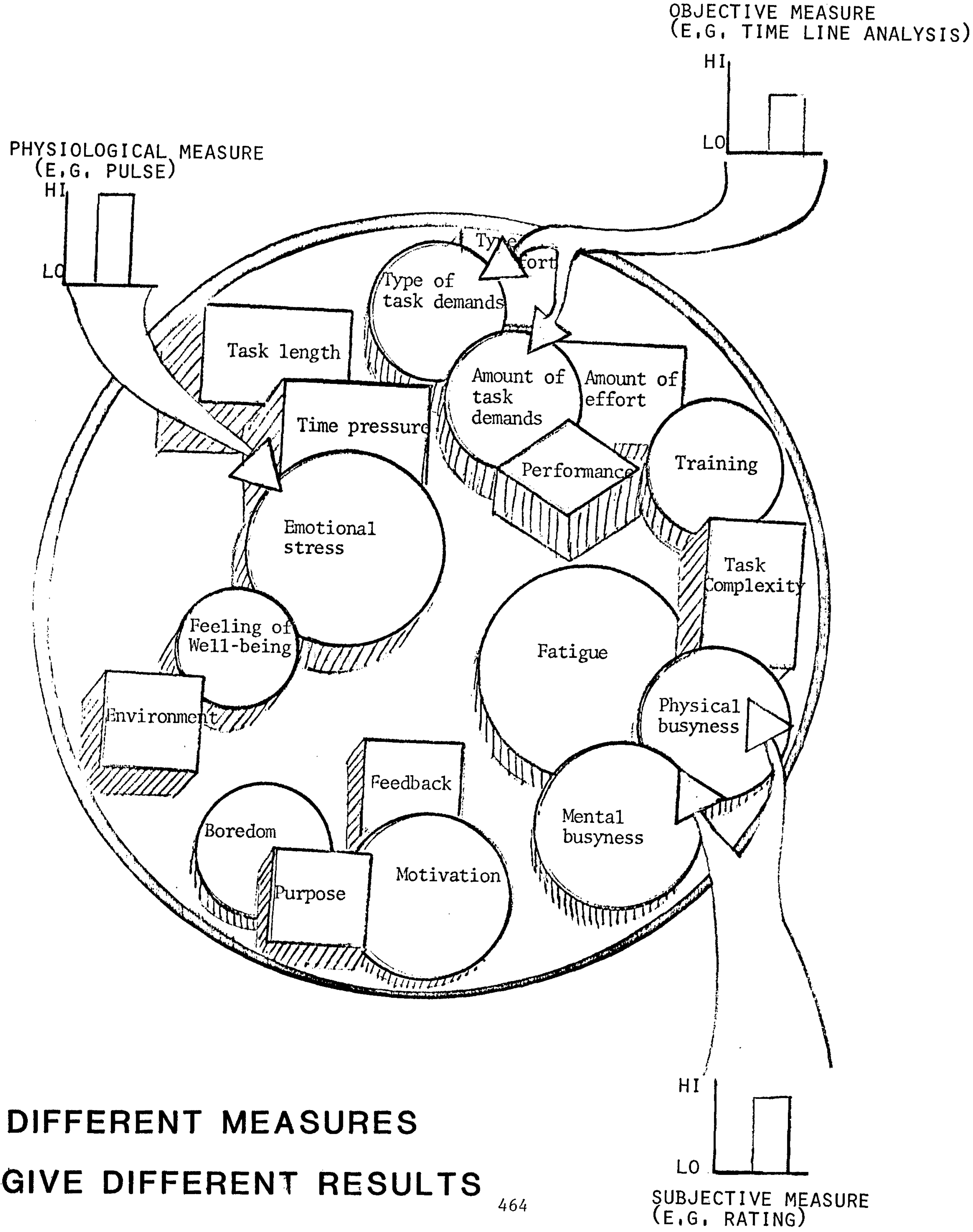


WORKLOAD IS RATED AS HIGH



IDENTICAL CIRCUMSTANCES CAN THUS RESULT IN DIFFERENT SUBJECTIVE EXPERIENCES

THE SIZE OF THE CIRCLE REPRESENTS THE MAGNITUDE OF AN INDIVIDUAL'S ESTIMATE OF WORKLOAD
THE SIZE OF THE SQUARES REPRESENTS THE "AMOUNT" THAT FACTOR PRESENT IN A GIVEN TASK
THE HEIGHT OF THE SQUARES REPRESENTS THE IMPORTANCE OR WEIGHT PLACED ON THAT FACTOR BY THE INDIVIDUAL



DIFFERENT MEASURES GIVE DIFFERENT RESULTS

ASSUMPTIONS

WORKLOAD IS:

☆ A SUBJECTIVE EXPERIENCE

☆ MULTI-DIMENSIONAL

☆ AMOUNT/NATURE OF COMPONENTS VARY

☆ COMPONENTS MAY COVARY

☆ IMPORTANCE OF COMPONENTS IS SUBJECTIVE & VARIABLE

☆ MEASURES ARE PROBES

☆ SUBJECTIVE EXPERIENCE AFFECTS "OBJECTIVE" MEASURES TOO

INITIAL RESEARCH GOALS:

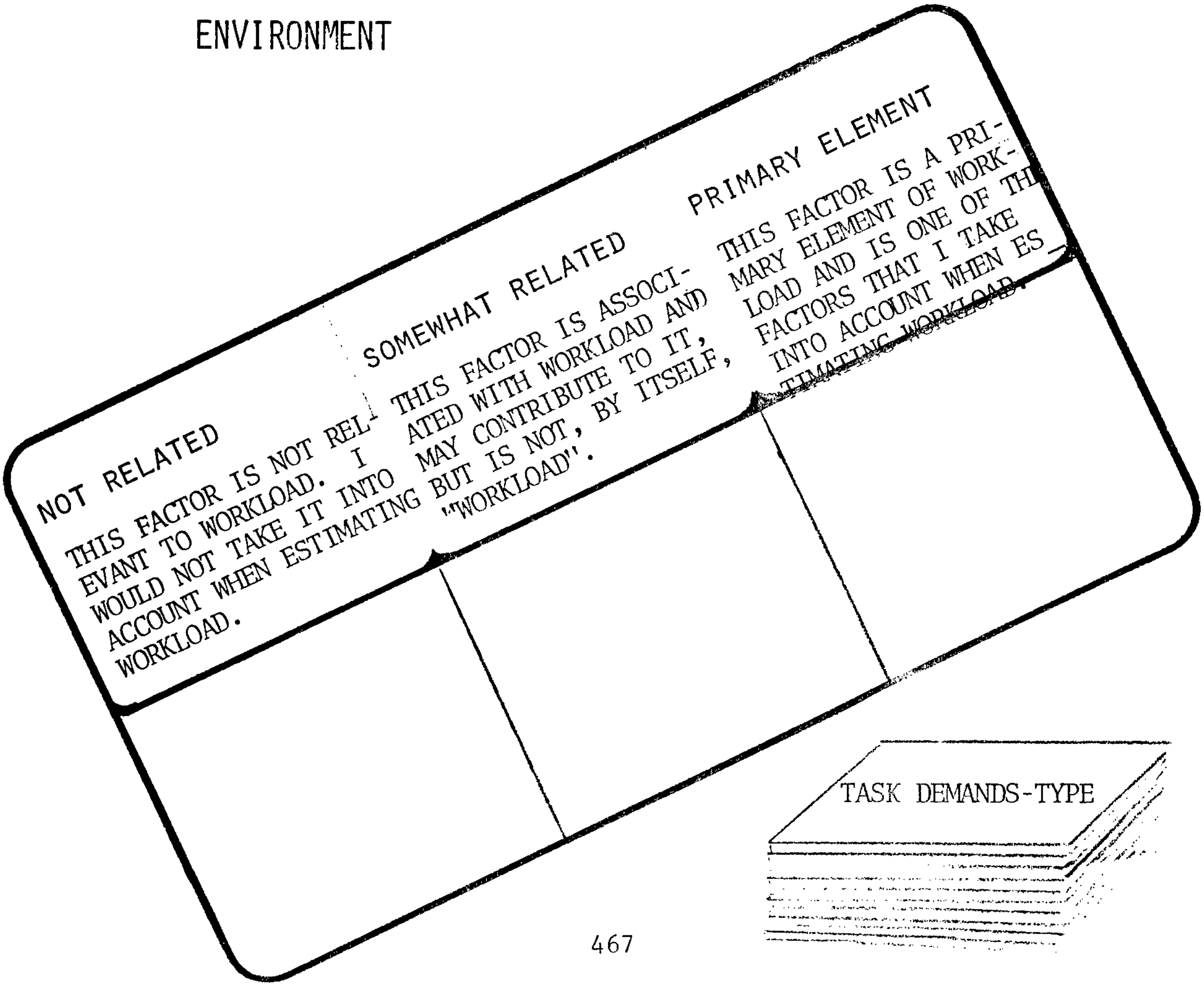
- IDENTIFY THE DIMENSIONS OF WORKLOAD
- MEASURE AND DESCRIBE INDIVIDUAL DIFFERENCES IN WEIGHTING THESE DIMENSIONS
- ESTABLISH A PRIORI ESTIMATE OF PRIMARY TASK LOAD IMPOSED BY SIMULATION SCENARIOS
- MEASURE SENSITIVITY OF COMMONLY USED WORKLOAD METRICS
- DEVELOP VALID, RELIABLE SUBJECTIVE RATING SCALES FOR WORKLOAD THAT REFLECT THE UNDERLYING FUNCTIONS
- IDENTIFY AND STANDARDIZE PRIMARY (E.G. IMBEDDED "SECONDARY") MEASURES OF WORKLOAD
- ANALYZE ON-LINE VERSUS RETROSPECTIVE ESTIMATES OF WORKLOAD
- INVESTIGATE PROBLEM OF MOMENTARY LOAD
- EXPLORE CIVIL AVIATION COMMUNICATIONS AS AN INDEX (AND CAUSE) OF WORKLOAD

METHOD FOR OBTAINING INFORMATION ABOUT THE IMPORTANCE OR WEIGHT PLACED ON DIFFERENT DIMENSIONS IN AN INDIVIDUAL'S CONCEPT OF WORKLOAD

DIMENSIONS TO BE RATED:

TASK DEMANDS - TYPE
TASK DEMANDS - AMOUNT
TASK COMPLEXITY
TIME PRESSURE
FEEDBACK
PURPOSE OF TASK
FATIGUE
EMOTIONAL STRESS
TASK LENGTH
ENVIRONMENT

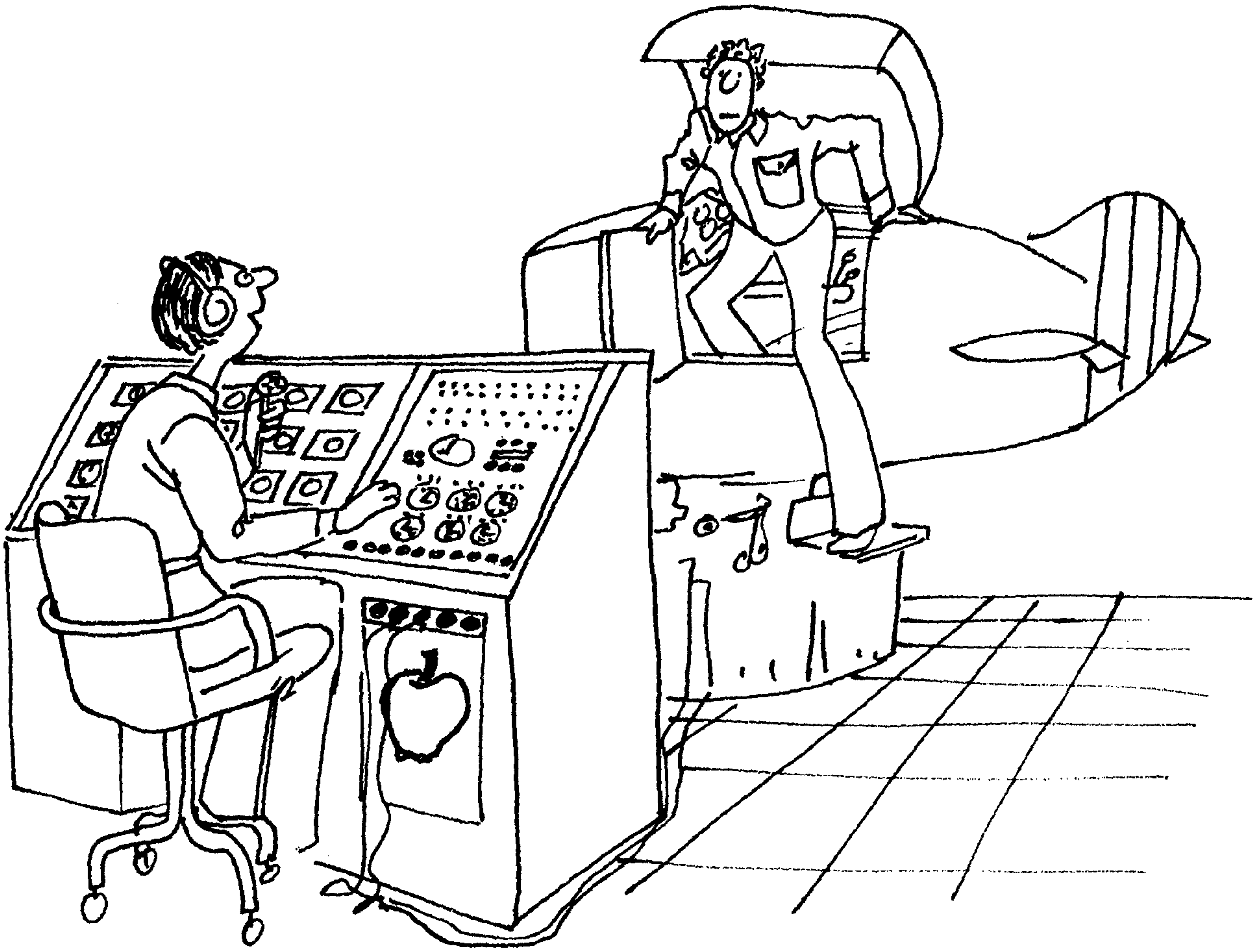
PERFORMANCE
BOREDOM
TRAINING
MENTAL BUSYNESS
PHYSICAL BUSYNESS
MOTIVATION
FEELING OF WELL-BEING
EFFORT - TYPE
EFFORT - AMOUNT



ON A SIMPLE TASK........MANY DIMENSIONS CAN BE VARIED........SO THAT MULTIPLE ASPECTS OF THE SUBJECTIVE EXPERIENCE OF WORKLOAD AND RELATED FACTORS CAN BE STUDIED

- DISCRIMINABILITY
- RATE
- ENVIRONMENT
- FEEDBACK
- COMPLEXITY
- PREDICTABILITY
- TRAINING
- PURPOSE
- LENGTH

- MENTAL EFFORT
- PHYSICAL EFFORT
- PERCEPTUAL EFFORT
- FATIGUE
- EMOTIONAL STRESS
- MOTIVATION
- FEELING OF WELL-BEING
- JOB SATISFACTION
- BUSYNESS



MOTION-BASE, GENERAL AVIATION SIMULATOR USED IN WORKLOAD ASSESSMENT RESEARCH AT AMES RESEARCH CENTER (LINK GAT-1)

BI-POLAR ADJECTIVE SCALES

OVERALL WORKLOAD	LO	HI
OWN PERFORMANCE	LOUSY	GREAT
TRAINING	TOO LITTLE	TOO MUCH
ATTENTION	CONSTANT	NONE
MOTIVATION	INDIFFERENT	EXCITED
STRESS LEVEL	RELAXED	VERY TENSE
ENERGY LEVEL	ENHANCED	REDUCED
PHYSICAL STATE	WIDE AWAKE	EXHAUSTED
TIME PRESSURE	NONE	VERY RUSHED
TASK DIFFICULTY	VERY EASY	VERY HARD
TASK COMPLEXITY	VERY SIMPLE	VERY COMPLEX
ACTIVITY LEVEL	IDLE	VERY BUSY
PHYSICAL EFFORT	LO	HI
MENTAL EFFORT	LO	HI
SENSORY EFFORT	LO	HI

DETERMINATION OF SENSITIVE MEASURES OF PILOT WORKLOAD AS A FUNCTION OF THE TYPE OF PILOTING TASK

WALTER W. WIERWILLE
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061

SUMMARY

The purpose of our present work, sponsored by NASA-AMES, is to examine the sensitivity, intrusion, and transferability of a variety of workload assessment techniques. The study will use four different simulated piloting tasks, emphasizing psychomotor, perceptual, mediational, and communications aspects. Pilot loading levels will be systematically adjusted. Our simulation facility is a GAT-1B that has been modified and instrumented for workload estimation techniques measurement. The flight simulator itself has three degrees of physical motion and a full complement of IFR instruments.

Recently we completed the experiment emphasizing the psychomotor aspect of flight. Instrument-rated pilots flew instrument approaches under three combined settings of the independent variable: increasing turbulance and decreasing longitudinal stability. Twenty different workload measures were taken between the outer and middle markers, only five of which showed statistically reliable changes as a function of the independent variable. Included in the five were: two rating scales, one measure of control movement activity, pulse rate, and one measure of time estimation. The results of the experiment are to some extent surprising, for they indicate that several "accepted" measures of workload are not reliably sensitive to the kinds of psychomotor load which pilots encounter.

We are currently planning the perceptual and mediational (cognitive) experiments, and we expect to have the results of these two experiments in mid-1982.

INTRODUCTION

The increasing complexity of aircraft systems and the changing roles of pilots and other aircrew personnel have resulted in the need for techniques to measure operator workload in a wide range of situations and tasks. One need only initiate a preliminary survey of the literature on operator workload assessment techniques to discover that a voluminous mass of information has accumulated rapidly in the past two decades. However, major reviews of this literature have concluded that while workload research has advanced in both scope and technology, basic questions remain to be answered for the practitioner who wishes to select workload measures for a given application (Wierwille and Williges, 1978). Hicks and Wierwille (1979) have pointed out that, in particular, the lack of information on the relative sensitivity, the degree of intrusion, and the range of transferability of individual techniques makes it difficult for a practitioner to select workload estimation techniques for a given task.

The purpose of our present work is to help fill the need for practical information. Specifically, techniques for measurement of pilot workload are being selected and compared to determine their relative sensitivity, intrusion, and transferability.

Before proceeding with further discussion and results of our experiments up to the present, it would be helpful to define the terms, sensitivity, intrusion, and transferability. Sensitivity can be defined as the relative ability of a workload estimation technique to discriminate statistically significant differences in operator loading. High sensitivity requires discriminable changes in the score means as a function of load level and low variation of the scores about the means. Intrusion can be defined as an undesirable change in the task for which workload is being estimated, resulting from the introduction of a workload measurement technique or apparatus. And, transferability is the relative ability of a workload estimation technique to remain sensitive when being applied in situations requiring different operator behaviors or skills.

Unfortunately, there has been no definitive major effort aimed at sensitivity, intrusion, and transferability. As a result, progress in determining which workload estimation techniques should be selected for a given application has been painfully slow. When asked which techniques are sensitive in a given piloting situation, an honest workload researcher has difficulty responding. The danger is that in a given application insensitive techniques may be used. These techniques would show no substantial change in workload when in fact there is a change. Unless one knows that a technique is sensitive in a given situation, one has no assurance the evaluation of workload in an experimental situation will result in definitive conclusions.

In our work we have arbitrarily divided piloting behavior into four categories: psychomotor, perceptual, mediational, and communications. These four behaviors are those suggested by Berliner, Angell, and Shearer (1964), in their list of universal operator behaviors (See also Wierwille and Williges, 1980). Clearly, other task taxonomy categories might have been chosen. However, the Berliner, et al categories do appear to reflect the major categories of behaviors exhibited by pilots and other aircrew members.

Our evaluation of workload estimation techniques in psychomotor tasks has been completed. Results of the study will be summarized in the following sections of this paper. More complete descriptions appear on Connor 1981 and Connor and Wierwille, 1982.

Presently, we are in the final stages of planning for the experiments emphasizing perceptual behavior and mediational behavior. In both cases pilots will fly the simulator in the simple task of maintaining heading, airspeed, and altitude. For the perceptual task they will also perform a forced-pace visual search task presented through the windscreen using an Ektagraphic display. The complexity of the search task will become the independent variable. For the mediational task, navigation problems will be presented. The problems will be forced-pace, but will not require computational aids. We expect to have the results of these two experiments in mid-1982. After these experiments have been completed we will also plan and carry out a simulated task involving communications. In all the experiments, loading level will be the independent variable, and technique scores will be the dependent variables. By conducting the experiments using this philosophy, we can obtain direct comparisons of the sensitivity of various techniques. Transferability will be evaluated by determining which (if any) techniques remain sensitive from one experiment to another . And, since primary task measures are taken for all techniques, intrusion can be determined by comparision of primary task measures with and without the other workload measurement techniques.

Clearly, the studies we are performing must necessarily be limited in scope, and they will not answer all important questions about sensitivity, intrusion, and transferability. Nevertheless we believe they will be very helpful to practitioners who must evaluate workload in realistic aircraft environments.

REVIEW OF THE EXPERIMENT ON PSYCHOMOTOR LOAD

Subjects

Six male instrument-rated pilots served as subjects in this experiment. The flight time of the subjects ranged from 500 to 2700 hours with a mean of 1300 hours.

Apparatus

The primary apparatus in this experiment was a modified flight task simulator (Singer Link, Inc., General Aviation Trainer, GAT-1B). The simulator had three degrees of freedom of motion (roll, pitch, and yaw). Translucent blinders were used to cover the windows of the simulator to reduce outside distractions and cues and to aid in the control of cockpit illumination.

Several modifications to the flight simulator were made for the experiment. These modifications permitted primary task load manipulation, secondary task operations, response measurement, and scoring. Primary task load manipulation was accomplished by changing aircraft pitch stability and random windgust disturbance level simultaneously. Three load conditions were developed: low, medium, and high, as shown in Table 1.

Table 2 provides a list of the workload measurement techniques selected for inclusion in the present study. These techniques were selected on one of two bases. First, evidence was found which indicated that the measures might be sensitive indicators of pilot workload in both simulated and operational flight. Second, previous research had shown that these measures could be useful in a variety of tasks relevant to the flight environment. A review of the twenty techniques selected can be found in Connor (1981).

Experimental Design

A complete 3 x 20 within-subject design was used for the sensitivity analysis. Load was the factor with three levels. Measurement technique (Table 2) was the factor with twenty levels.

Workload measures from different techniques were taken simultaneously on some of the data collection runs. Only those measures which were not likely to affect each other were taken simultaneously. Table 3 shows the scheme used for combining different measurement techniques for data collection. The combination of measurement techniques shown in the table was, to an extent, based on previous investigations of workload. Hicks and Wierwille's (1979) study supported the combination in condition 2. The two rating scales were administered in separate measurement conditions to prevent the ratings on one scale from biasing the ratings on the other scale. The secondary task measures were divided among several conditions because of potential intrusion and interference. Vocal measures were recorded from the two secondary tasks which required a verbal response as per Schiflett and Loikith's (1979) recommendation.

It should be noted that primary task measures were recorded on all subjects and on all data collection flights for the intrusion analysis. However, only data from measurement condition 1 were used for the sensitivity analysis of the primary task measures.

The intrusion analysis was designed to examine the effect of measurement condition, and the interaction of measurement condition with load on primary task performance. Data for all primary task measures were therefore collected for each flight performed in the six measurement conditions.

General Procedure

After receiving instructions, subjects flew nine familiarization flights in the simulator. These flights were similar, but not the same as, the data collection flights. All subjects flew the familiarization flights in the same order. Steady crosswinds were introduced for each run, and subjects were given heading corrections.

After the familiarization session, the subjects participated in three data collection sessions. The familiarization session and each data collection session were held on a different day.

Each data collection session consisted of two sets of a warm-up practice flight and three data collection flights. The practice flight was the same as the first data collection flight. Since the data collection flights were counterbalanced, equal amounts of practice were provided for the low, medium, and high load conditions. The data collection flights also contained steady crosswind conditions, for which the subject was given heading corrections. The purpose of introducing steady crosswinds was to disguise the load conditions, thereby requiring subjects to fly each flight as a separate entity.

Flight Task Procedures

The flight task in this experiment was an Instrument Landing Systems (ILS) approach to the Seaport Beach runway (29L) which is instrumented in the Singer Link GAT-1B aircraft simulator. Prior to the beginning of a flight, the simulated aircraft was positioned 5 miles outbound from the Seaport Beach outer marker on the 108 degree radial, heading into the wind. When ready to begin, the experimenter informed the subject of the wind direction and speed, and gave him a heading correction for the crosswind. When contacted by the experimenter, the subject took off and climbed to 2000 feet. The subject then flew directly to the outer marker by following the localizer at 100 miles per hour until the glide slope was intercepted. Upon interception of the glide slope, the subject reduced airspeed to 80 miles per hour and proceeded down the glide slope while following the localizer to a landing. Data were recorded between the outer and middle markers. For the opinion measures, subjects gave ratings for the flight segment between the outer and middle markers immediately after landing and parking the simulated aircraft.

Results

Sensitivity Analysis

The computed scores for each technique were first converted to Z-scores (normalized scores) so that technique measure units would not affect the sensitivity analysis. Subsequently, an overall analysis of variance was performed on the scores. Since Z-scores were used, a technique main effect was not possible. A significant main effect of load was found, $F(2,10) = 5.34$, $p < 0.0001$, and a significant load by technique interaction was found, $F(38,190) = 2.76$, $p \leq 0.05$.

The load by technique interaction indicated that the measurement techniques were differentially sensitive to load. Therefore, individual ANOVAs were used to isolate the sensitive techniques.

The individual ANOVAs indicated that five of the twenty measures were sensitive. They were the Cooper-Harper scale $F(2,10) = 16.39$, $p = 0.0007$; the Workload Compensation-Interference/Technical Effectiveness (WCI/TE) scale, $F(2, 10) = 31.15$, $p < 0.0001$; the time estimation standard deviation, $F(2,10) = 5.69$,

$p = 0.022$; the pulse rate mean, F (2,10) = 8.89, $p = 0.006$; and the control movements measure, F (2,10) = 33.84, $p < 0.0001$. The normalized means for each technique are plotted in Figures 1 through 5 as a function of load.

Newman-Keuls comparisons were then performed on the normalized means of the sensitive measures. The comparisons included low vs. medium, medium vs. high, and low vs. high load conditions. Results indicated that all differences were significant at $p < 0.05$, except for pulse-rate mean (low vs. medium and medium vs. high) and time estimation standard deviation (low vs. high).

A logical classification of techniques based on demonstrated sensitivity was generated from an examination of the Newman-Keuls comparisons, as shown in Table 4. Techniques which demonstrated sensitivity to all pairs of load conditions (i.e., low vs. medium, medium vs. high, and low vs. high) were included in class I. These measures are preferred over other techniques which demonstrated only partial sensitivity, or no sensitivity in the present study. Techniques which showed sensitivity to some differences in load conditions (but not all) were included in class II. These measures are less preferred than class I techniques, but are more preferred than class III techniques. Class III techniques did not demonstrate sensitivity to load in the present study. This class includes all techniques except those in class I and class II.

One possible reason that only five of the twenty workload assessment techniques demonstrated sensitivity in the present study is that the other techniques simply required a greater number of subjects to show a significant effect of load. It is possible to estimate the sample size required to detect a reliable load effect for a given workload assessment technique at specified levels of significance and power. These calculations were performed for those techniques which did not demonstrate sensitivity in the present study, to provide an indication of the practical costs of achieving statistical significance. The procedure used for estimating the sample size required for finding sensitivity is described by Bowker and Lieberman (1959). Sample sizes were estimated for a significance level of 0.05 and for a power of approximately 0.80. The results of these estimates are presented in Table 5.

Intrusion Analysis

The equipment and procedures used for some workload assessment techniques may interfere with performance on the primary (flight) task. In the present experiment, data for the twenty measurement techniques were recorded in six measurement conditions as shown in Table 3. These six measurement conditions differed in the equipment and procedures used for data collection. The purpose of the intrusion analysis was to examine the effect of these measurement conditions on primary task performance.

The equipment and procedures used in measurement condition 1 were assumed to be unobtrusive to primary task performance. Primary task performance in this condition was therefore used as a standard of comparison for primary task performance on the other five measurement conditions. The measures of primary task performance which were used for these comparisons included scores on localizer rms error, glide slope rms error, and control movements per second.

A multivariate analysis of variance (MANOVA) was performed to examine the effect of condition, load, and the interaction of condition and load on the primary task measures. Only the main effect of load was found to be significant F (2,10) = 9.42, p = 0.0002. Because there was no significant effect of condition nor significant interaction of condition with load, it can be concluded that the physiological measuring equipment and the secondary tasks did not significantly affect pilot performance in terms of the three primary task variables.

Conclusions

This study has shown that five measures of workload estimation were sensitive indicators of load in a piloting task that is predominantly psychomotor in nature. Another fifteen measures, believed to be "good" measures of workload, showed no reliable effect. The main conclusion that must be drawn from the study is that few measures are sensitive to psychomotor load.

Of the five techniques demonstrating sensitivity, only three exhibited monotonic score increases with load as well as statistically reliable differences between all pairs of load levels. Consequently, only the three meet all criteria for sensitivity to psychomotor load. These class I techniques are the ones that are recommended for measurement of psychomotor load:

Cooper/Harper ratings,
WCI/TE ratings, and
Control movements per second.

The other two techniques showed sensitivity to psychomotor load, but did not discriminate between all pairs of load levels. These class II techniques are:

Time estimation standard deviation, and
Pulse rate mean.

These measures would be helpful in evaluating psychomotor load, but they should not be relied on exclusively. At least one class I technique should also be used in conjunction with these measures.

It is worth noting that only two opinion measures were taken in the present experiment, and both proved sensitive. This suggests that well-designed rating scales are among the best of techniques for evaluating psychomotor load. In regard to the primary task measures, the control movements measure alone was sensitive. However, this measure is also the only primary task measure which reflected "strategy" of the pilot. Consequently, one could speculate that selecting a primary task measure that reflects strategy will most likely result in good sensitivity.

Fifteen (techniques) measures showed no reliable change as a function of load. When these fifteen measures were subjected to a power analysis to determine sample size, the number of subjects required ranged from 12 to well over 100 (Table 5). One can only conclude that at best the fifteen measures, as taken, are much less sensitive to psychomotor load than the five appearing in Classes I and II. Of course, there is always the possibility that the measures would be sensitive to loading along other dimensions of human performance, such as psychomotor tasks of a different nature, or mediational or cognitive tasks, for example.

In regard to intrusion, this experiment showed that no significant interference occurred for the physiological measures or for the secondary task measures. Performance as measured using three primary (flight) task measures showed no reliable changes as a function of addition of these measures.

In general, the results of the experiment show that there are wide variations in the sensitivity of workload estimation measures. Great care must be taken in selecting measures for a given experiment. Otherwise, it is possible that no changes in workload will be found, when indeed there are changes.

References

Berliner, C., Angell, D., and Shearer, D. J. Behaviors, measures, and instruments for performance evaluation in simulated environments. Paper presented at the Symposium and Workshop on the Quantification of Human Performance, Albuquerque, New Mexico, August, 1964.

Bowker, A. H. and Lieberman, G. J. Engineering statistics. New Jersey: Prentice-Hall, Inc., 1959.

Connor, S. A. A comparison of pilot workload assessment techniques using a psychomotor task in a moving base aircraft simulator. Master's Thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, October, 1981.

Connor, S. A. and Wierwille, W. W. Comparative evaluation of twenty pilot workload assessment measures using a psychomotor task in a moving base simulator. Moffett Field, CA: NASA-Ames Research Center, (Forthcoming report).

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

Schiflett, S. G. and Loikith, G. J. Voice stress as a measure of operator workload. Patuxent River, Maryland: Naval Air Test Center, Technical Memorandom TM 79-3 SY, December 31, 1979.

Wierwille, W. W. and Williges, R. C. Survey and analysis of operator workload assessment techniques. Blacksburg, Virginia: Systemetrics, Inc. Report No. S-78-101, September, 1978.

Wierwille, W. W. and Williges, B. H. An annotated bibliography on operator mental workload assessment. Patuxent River, Maryland: Naval Air Test Center Report No. SY-27R-80, March, 1980.

Acknowledgements

The author wishes to thank Mr. Sidney A. Connor, who carried out the psychomotor workload experiment, and Mrs. Sandra Hart, NASA-Ames Research Center, for helpful technical suggestions.

TABLE 1

Primary Task Load Conditions

	LOAD CONDITION		
	Low	Medium	High
RANDOM GUST LEVEL	Low	Medium	High
Estimated Std. Dev. (mph)	0	2.7	5.9
PITCH STABILITY	High	Medium	Low
a. Control input to pitch rate output equivalent gain (degrees/s per % of control range)	0.522	3.560	7.83
b. Control input to pitch rate output equivalent time constant (s)	0.097	0.660	1.45

TABLE 2

Workload Assessment Techniques Which Were Tested in the Present Experiment

---

OPINION

1. Cooper-Harper Scale
2. WCI/TE Scale

SPARE MENTAL CAPACITY

3. Digit Shadowing (% errors)
4. Memory Scanning (Mean time)
5. Mental Arithmetic (% errors)
6. Time Estimation Mean (Seconds)
7. Time Estimation Standard Deviation (Seconds)
8. Time Estimation Absolute Error (Seconds)
9. Time Estimation RMS error (Seconds)

PHYSIOLOGICAL

10. Pulse Rate Mean (Pulses per minute)
11. Pulse Rate Variability (Pulses per minute)
12. Respiration Rate (Breath cycles per minute)
13. Pupil Diameter (Normalized units)
14. Voice Pattern (Digit Shadowing Task)
15. Voice Pattern (Mental Arithmetic Task)

EYE BEHAVIOR

16. Eye Transition Frequency (Transitions per minute)
17. Eye Blink Frequency (Blinks per minute)

PRIMARY TASK

18. Localizer RMS Angular Position Error (Degrees)
19. Glide Slope RMS Angular Position Error (Degrees)
20. Control Movements per second
    (Aileron + Elevator + Rudder)

TABLE 3

Combination of Measurement Techniques
for Data Collection

Measurement Condition	Measurement Techniques
1.	Cooper-Harper Scale Pupil Diameter Eye Transition Frequency Eye Blink Frequency Localizer RMS Error Glide Slope RMS Error Control Movements
2.	WCI/TE Scale Pulse Rate Mean Pulse Rate Variability Respiration Rate
3.	Digit Shadowing Voice Pattern
4.	Memory Scanning
5.	Mental Arithmetic Voice Pattern
6.	Time Estimation (Mean) (Std. Dev.) (Abs. Error) (RMS Error)

TABLE 4

Logical Classification of Techniques Based on Demonstrated Sensitivity

---

Class I: Complete Sensitivity Demonstrated
  Cooper-Harper Scale
  WCI/TE Scale
  Control Movements/Unit Time

Class II: Some Sensitivity Demonstrated
  Time Estimation Standard Deviation*
  Pulse Rate Mean **

Class III: Sensitivity Not Demonstrated
  All Other Techniques (See Table 5)

---

*Double valued function
**Limited sensitivity

TABLE 5

Estimated Sample Sizes Required for Achieving a Significant Load Effect for Techniques not Demonstrating Sensitivity

Technique	Estimated Sample Size
SPARE MENTAL CAPACITY:	
Digit Shadowing	18
Memory Scanning	>100
Mental Arithmetic	25
Time Estimation (Mean)	53
Time Estimation (Abs. Error)	>100
Time Estimation (RMS Error)	85
PHYSIOLOGICAL	
Pulse Rate Variability	45
Respiration Rate	15
Pupil Diameter	>100
Speech Pattern (D. Shadow.)	28
Speech Pattern (M. Arith.)	>100
EYE BEHAVIOR	
Eye Transition Frequency	42
Eye Blink Frequency	25
PRIMARY TASK	
Localizer RMS Error	12
Glide Slope RMS Error	41

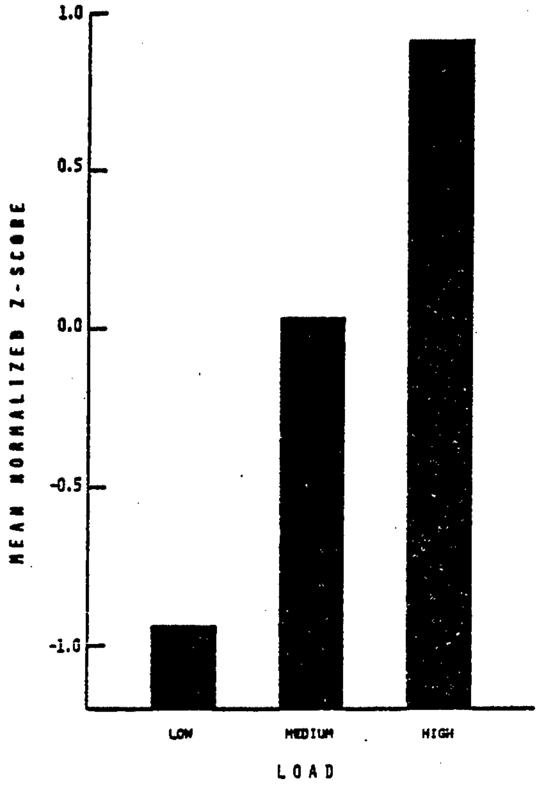


Figure 1. Mean normalized scores for the Cooper-Harper rating scale measure plotted as a function of load.

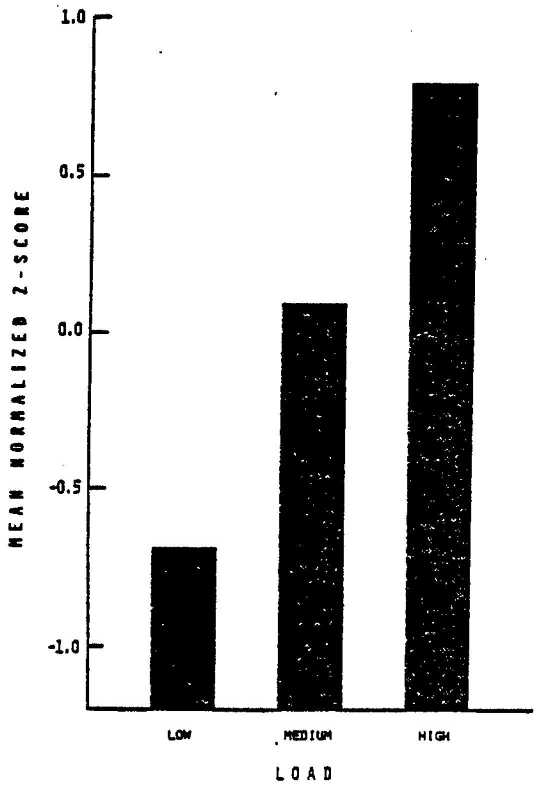


Figure 2. Mean normalized scores for the WCI/TE rating scale measure plotted as a function of load.

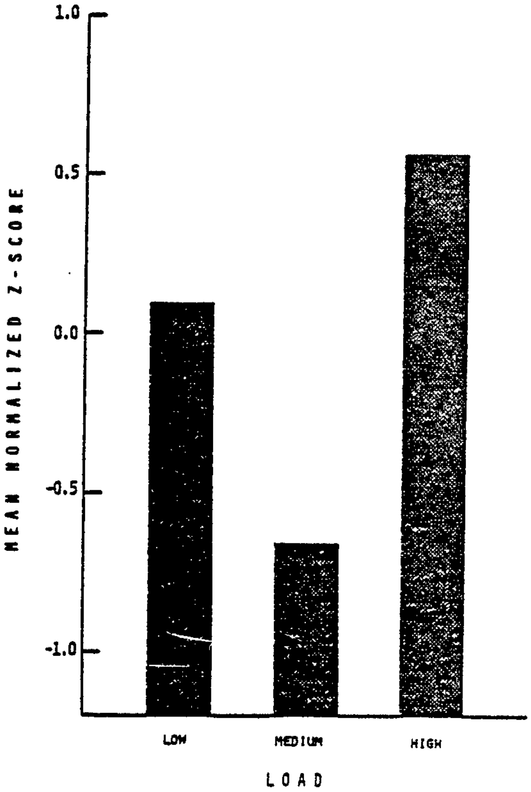


Figure 3. Mean normalized scores for the time estimation standard deviation measure plotted as a function of load.

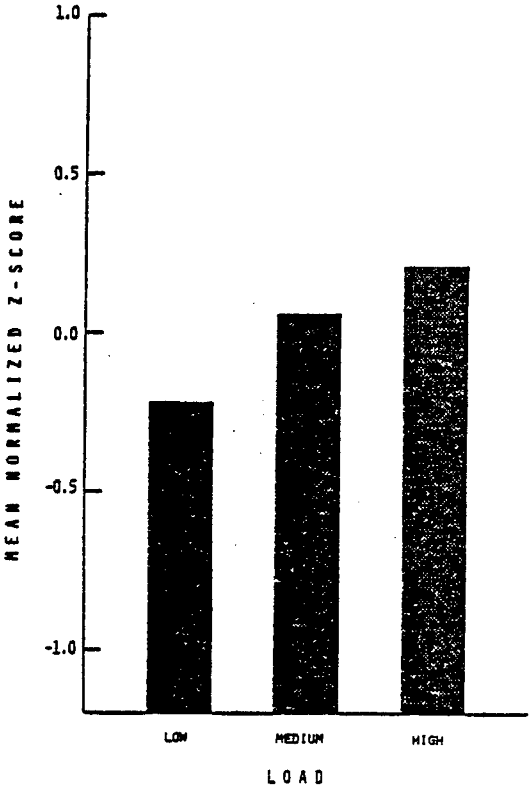


Figure 4. Mean normalized scores for the pulse rate mean measure plotted as a function of load.

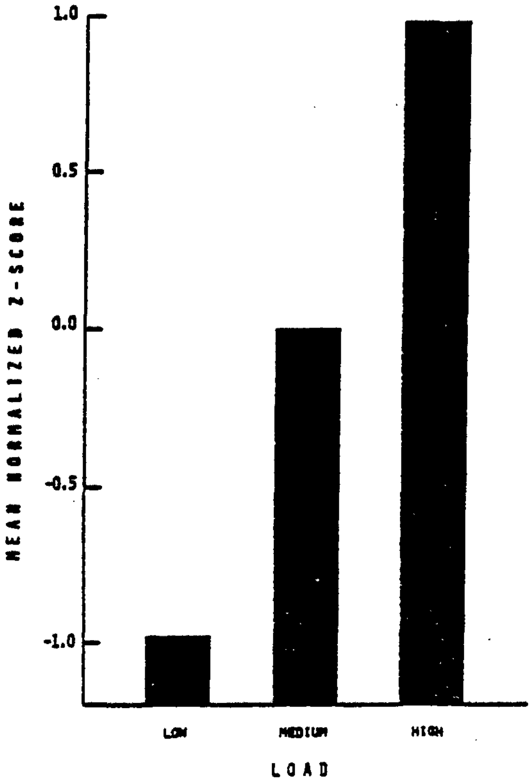


Figure 5. Mean normalized scores for the control movements measure plotted as a function of load.

SECTION III

TAILORING FLIGHT TESTS TO IDENTIFY PILOT DYNAMICS AND TASK PERFORMANCE

An Annotated Bibliography of Objective Pilot Performance Measures

by

T. R. Mixon & W. F. Moroney

Naval Postgraduate School

Monterey, CA

INTRODUCTION

The measurement of pilot performance is a "sine qua non." Without measurement one cannot discriminate success from failure, progression from regression. Measurement is a means for determining where one is, where one has been, and potentially where one is going. S.S. Stevens defined measurement as "the assignment of numerals to objects or events according to rules." This document reviews attempts to assign numerals to pilot activities according to rules. Thus, the emphasis is on objective performance measures, i.e., those measures which can be obtained without observer-system interaction.

Some of the areas on which pilot performance measurement (PPM) impacts are listed in Figure one. Workload is only one of those areas, yet to some extent "Pilot Workload Measurement" could have been inserted at the center of the figure. Indeed pilot performance and workload are closely interviewed, both have been measured and manipulated under a variety of circumstances ranging from mathematical model to the real world. As suggested on Figure 2, these are advantages and disadvantages associated with gathering data under each circumstance.

Buckout's review in 1962 was the last comprehensive examination of the pilot performance measurement literature. The annotated bibliography discused in this paper attempted to: (1) gather the PPM literature written subsequent to 1962 into one source (2) describe the scenarios and measures used in collecting PPM data; and (3) summarize the major premises and findings of each article.

A variety of sources including computer based literature searchs were used to identify candidate articles. Ultimately all referenced material was divided into three catergories: (1) objective pilot performance measures; (2) subjective pilot performance measures; and (3) general analysis and review articles. Each of these sections are discussed separately below:

I OBJECTIVE PILOT PERFORMANCE MEASURES

A total of 189 different reports which use objective pilot performance measures were located. For each report the following data were provided:

1. Standard Bibliographics Reference Material: Author(s), Title, Source(s)
2. Subjects. Number and type of personnel used in the research.
3. Equipment. Each experiment utilized either a laboratory, a flight simulator, or an actual aircraft. The distinguishing line between "laboratory" and "flight simulator" was not always clear; usually if a full-sized aircraft cockpit was employed, this was classified as a "simulation" experiment. When more than one form of equipment was utilized, this was so indicated.
4. Scenario. A broad definition of airborne flight including missions and flight segments.
5. Measures. Observed parameters or variables measured in the objective sense without human perceptions or judgements. These generally fell into six classes; (1) physiological, (2) aircraft systems, (3) man-machine system, (4) time, (5) frequency, and (6) combined measures. Mathematical and statistical transformations applied to each measure are included in parentheses after the measure.
6. Summary. A capsulized synopsis of the purpose of the experiment, experimental conditions, and the results. Brevity was preferred over repetitious statements of facts. No attempt was made to review or critique a document.

Figure 3 is a sample of the reporting format. The articles were classified according to the following format:

Descriptor		Article Number
Field Conditions	Fixed Wing	100-120
	Rotary Wing/VSTOL	120-149
Simulator Conditions	Fixed Wing	200-276
	Rotary Wing/VSTOL	277-287
Laboratory Conditions		300-319
Combination of Field Conditions, Simulator and/or Laboratory Conditions		400-434

After all the literature had been reviewed 188 different measures were identified. Matrices, which divided the 188 performance measures into six main classes: physiological, aircraft systems, man-machine systems, time, frequency, and measures of effectiveness/other were developed. By using these matrices, the researcher can identify articles of interests which use a common performance measure.

II SUBJECTIVE PILOT PERFORMANCE MEASURES

A small sample (N=34) of subjective pilot performance measurement studies were included to contrast the pilot rating method with the objective measures reported in Articles 100-434. In addition to the bibliographic data this section contains either the author's original abstract, modified slightly in some cases, or a modification of the sources' introductory material.

III RELATED ANALYSIS AND REVIEWS

This section contains articles which while related to pilot performance measurement did not contain objective or subjective data but rather, addressed related issues such as data analysis, simulation, transfer of training, prediction of pilot performance, math models, etc. In addition to the bibliographic data, this section contains either the author's original abstract, modified slightly in some cases, or a modification of the source's introductory material.

INDEXES

In an effort of this size retrieval of the article(s) of interest is critical. Therefore, in addition to the matrices described previously the following indexes were developed: Author, Subject Matter, Scenario, Performance Measurement, Source, and Accession Number.

REPORT AVAILABILITY

The annotated bibliography described above will be released by the Naval Training Equipment Center as IH 330 in March of 1982. A limited distribution annotated bibliography, based on a survey of the limited distribution and Department of Defense classified literature will be released by the Naval Postgraduate School in the Fall of 1982 as NPS 55-81-010PR.

The authors hope that these documents will (1) provide a means for integrating the PPM literature, and (2) serve as an impetus to develop a systematic approach to PPM.

ADDITIONAL MATERIAL

Readers who are interested in PPM as a means of evaluating skilled operator performance might wish to consult Mixon (1981).

REFERENCES

Buckout, Robert, A Bibliography on Aircrew Proficiency Measurement, Aerospace Medical Divison, Wright-Patterson AFB, OH, MRL-TDR-62-49, 1962, AD 283545.

Chapanis, A and Van Cott, H.P. Human Engineering Test and Evaluation in Van Cott H.P. & Kinkade, R.C. Human Engineering Guide to Equipment Design, Washington, D.C., Government Printing Office, 1972.

Mixon, Ted R., A Model to Measure Bombadier/Navigator Performance During Radar Navigation in Device 2F114, A6E Weapon System Trainer. Naval Postgraduate School, Monterey, CA, 1981, AD A098776.

Stevens, S.S., ed., Handbook of Experimental Psychology. New York: Wiley, 1951.

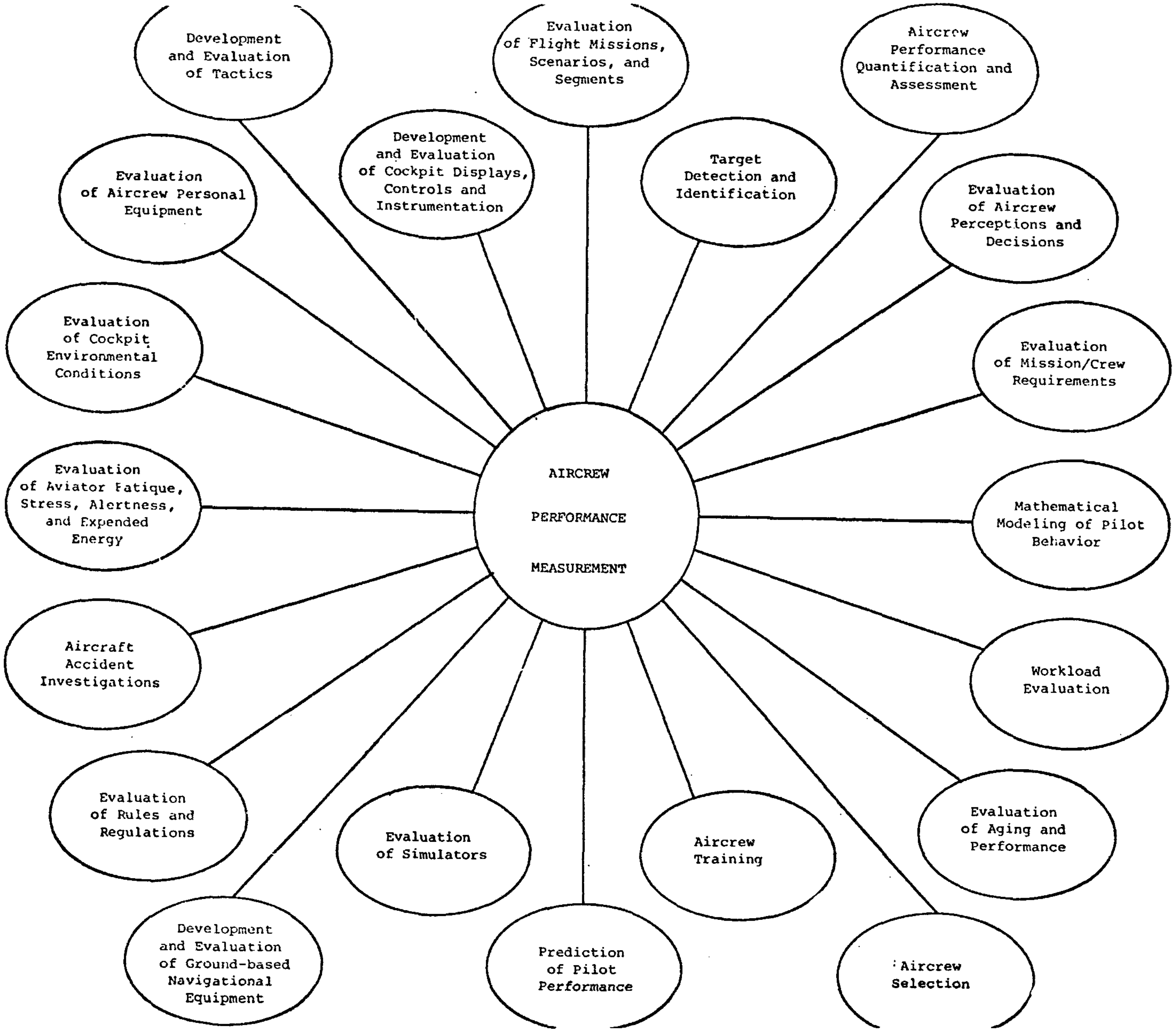


Figure 1

Areas on which Aircrew Performance Measurement Impacts

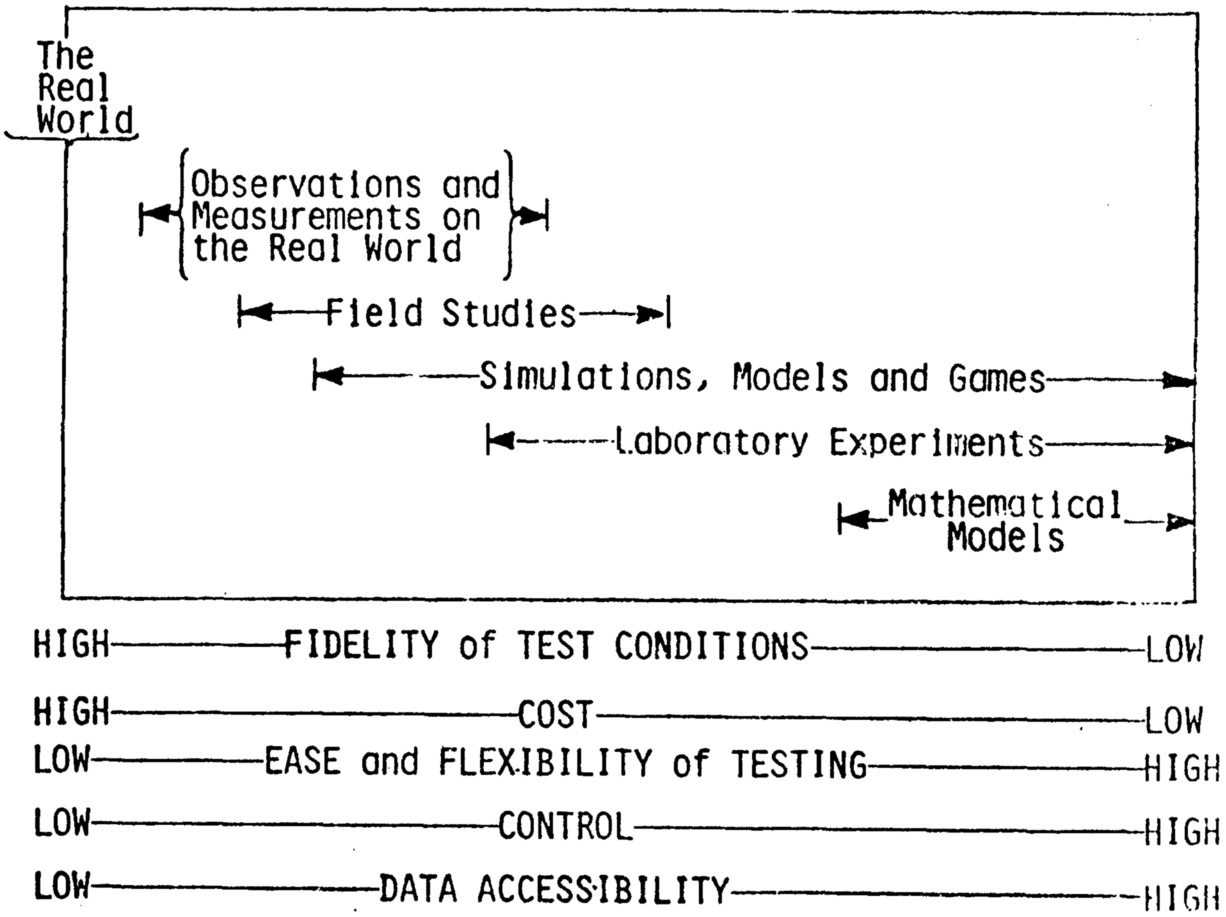


Figure 2

Conditions under which PPM and Workload data have been gathered and some associated Trade-offs (Modified from Chapanis & Van Cott, 1972)

266. KELLY, Michael, J., Wooldridge, Lee, Hennessy, Robert T., Vreuls, Donald, Barnebey, Steve F., Cotton, John C., and Reed, John C., Air Combat Maneuvering Performance Measurement, Canyon Research Group, Inc., Westlake Village, CA 91361, Contract No. F33615-77-C-0079, sponsored by Naval Training Equipment Center, Orlando, FL 32813, and Air Force Human Resources Laboratory, Brooks Air Force Base, TX 78235, NAVTRAEQUIPCEN IH-315/AFHRC-TR-79-3, September 1979, 142 pp. See also Proceedings of the 23rd Annual Meeting of the Human Factors Society, 1979, p. 324-328.

SUBJECTS:	30 fighter pilots.
EQUIPMENT:	Simulator for Air-to-Air Combat (SAAC) configured as an F-4 aircraft.
SCENARIO:	One-vs-one air combat maneuvering.
MEASURES:	Altitude rate (mean), opponent out of view (percentage of time opponent out of pilot's field of view), airspeed (mean), speedbrake (mean deflection), fuel flow (mean), relative altitude use (ratio of altitude standard deviations), energy management index (function of remaining fuel, fuel flow, airspeed and altitude), offensive time (sight angle less than 60 degrees), offensive time with advantage, throttle percentage time (idle, LO MIL, HI MIL, and afterburner), heading (root-mean-square and absolute average), lead time, time within range, roll rate, maneuvering rate (roll rate times altitude rate), ACM plane of action (composite of X, Y and Z), defensive time, angle of attack (percentage of time greater than 28 units), and aircraft kills (percentage of engagements ending in an AIM-9 success, gun success, ground impact, over-g or fuel exhaustion).
SUMMARY:	The goal of this study was to develop a preliminary measurement structure and measurement set for an automated Air Combat Maneuvering Performance Measurement (ACMPM) system which could be implemented on the SAAC. The measurement system was to provide valid and diagnostic performance information in real time. Using multivariate analysis, a measurement model containing 13 variables accounted for 51 percent of the performance variance and was able to discriminate between pilots of high and low skill with an accuracy of 92.1 percent. It was recommended that further analyses, developmental testing, and

Figure 3

Sample of reports included in annotated bibliography

A Study of Task Difficulty With A Subjective Rating Scale

D. W. Repperger, D. B. Rogers, R. E. Van Patten, J. Frazier

Air Force Aerospace Medical Research Laboratory
Wright Patterson Air Force Base, Ohio 45433

SUMMARY

The results of two experiments are discussed which relate to task difficulty and the effects of environmental stress on tracking performance. The first experiment involved 5 different sum of sine tracking tasks which humans tracked both in a static condition and under a 5 Gz acceleration stress condition. The tasks were designed in such a manner as to investigate workload measures and to compare our hypothetical design to subjective evaluations. The tasks were required to satisfy 5 criteria specified in mathematical terms.

The second experiment involved similar environmental stress conditions but in this case the tasks were constructed from deterministic functions with specially designed velocity and acceleration profiles. In both parts of this experiment, subjective evaluations were obtained and compared to the assumption that difficulty is related to magnitudes of velocity and acceleration profiles of the target tracking task. Phase Plane performance analysis was conducted across 7 subjects to study potential measures of workload or tracking difficulty.

INTRODUCTION

In the study of manual control theory, the systematic characterization of task difficulty has been a problem of considerable interest for many years. An extensive amount of work has been done in this area and a variety of studies indicating different measures related to workload are available in the Human Factors and Psychological literature. In the engineering literature, the classical paper by Cooper (reference 1) illustrates the motivation for such a characterization of task difficulty - a subjective rating scale for human tracking. The extent at which this subjective rating scale can be used to elict pilot response is best illustrated in reference 2 where a thorough study has been done to investigate and pinpoint the exact cause-effect relationships between pilot subjective ratings and handling qualities of aircraft. This study uses a decision tree type of analysis procedure to investigate the responses.

At the Air Force Aerospace Medical Research Laboratory, it is of interest in our research program to develop standard tasks or levels of tracking difficulty and to be able to estimate levels of difficulty associated with human tracking. Once a consistent set of tasks are developed which provide a basis or standard for tracking behavior, it is then possible to more closely evaluate the effects of stress on human tracking performance. The criteria for the design of the tracking tasks must be such that each task is required to be a sensitive indicator of performance change (between each task number) and, in addition, the requirement is made that the task is to be sufficiently sensitive as to show a performance decrement between the stress-non stress condition.

This study consisted of two separate experiments. Both parts of this investigation involved human tracking for target forcing functions with different acceleration and velocity profiles. It was desired to study a critical task concept (reference 3) based on a hierarchy of difficulty associated with the different target forcing functions. This approach differs from the classical critical task concept considered by Jex, et al. (reference 4) in which the controlled element would have dynamics that change. In our studies, the controlled element (figure (1)) remained the same; the tracking tasks varied in levels of difficulty based on our hypothesis of different velocity and acceleration profiles associated with each target forcing function. The motivation for this work was due to an interesting paper by Verplank (reference 5) in which he equated difficulty and stress in studying human response behavior within a vigilance paradigm.

SYMBOLS

$f(t)$ = The Target Forcing Function Signal
$e(t)$ = The Closed Loop Error Signal
$x(t)$ = The Output of The Plant (Controlled Element)
R = Radius in the $\dot{f}$ versus $\ddot{f}$ plane = $(\dot{f})^2 + (\ddot{f})^2$
μ = Median of the distribution of the error window histogram
e_{RMS} = Root Mean Square error score
X = mean e_{RMS} value
σ = standard deviation of e_{RMS} value
X_d = The deviation of a difference from the mean of the differences.
M_d = Mean of the n differences of paired observations.
$\underline{p}$ = Probability
$\overline{t}$ = t test statistic
t = time

METHOD

Subjects - Seven male United States Air Force volunteers participated in this experiment. They had prior training in both the G type of stress exposures and manual tracking tasks.

Design of The Target Tracking Task - Part I

The objective of this study was to develop the tracking tasks of different levels of difficulty and to study their ability to produce performance decrements between tasks (for a given experimental condition) and between experimental conditions (for the same task). For the first part of this study, it was decided to design five different tasks with the following constraints:
(1) Each tracking task will be zero mean, constant variance, sum of sines.
(2) Each forcing function when presented as a replication will have a random

initial phase angle for each frequency component. The phase angle must be a prime multiple of the fundamental frequency and not a linear constant multiple of any other frequency component.

(3) Each forcing function will have a random phase angle between each frequency.

(4) Due to human physiological exposure limits in the design of the acceleration experiment, the length of each task was set at 15 seconds.

(5) The amplitudes of all the sinusoids are scaled so such that they all have equal power and produce the same displacement on the CRT (display). The open loop and autopilot runs of this study which verify this fact are presented in the sequel.

(6) The component frequencies of the sinuosids are "relatively prime" multiples of a fundamental frequency.

(7) A shift in frequency content is required so that $ff_i > ff_j$ is true if $i>j$, $i,j=1,2,..5$ where the frequency content of ff_i is higher than that of ff_j. The procedure for obtaining this desired result is discussed in reference 6.

Using a measure of difficulty denoted as R (the distance in the target phase plane (figure (2))) where R satisfies:

$$R^2 = (\dot{f})^2 + (\ddot{f})^2 \qquad (1)$$

Then table I illustrates the values of R obtained for the 5 different tasks chosen in Part I of this study.

Table I Results of The Open Loop and Autopilot Simulations

Forcing Function (or Task) Number	Open Loop Error RMS * 2351.	Autopilot Error RMS * 2351	R^2(mean) For The Forcing Function	R^2(s.d.) For The Forcing Function
#1	718.6	477.8	0.352	0.290
#2	718.6	477.8	0.662	0.611
#3	718.6	477.8	1.212	1.140
#4	718.6	477.8	2.151	1.897
#5	718.6	477.8	3.509	3.2309

In this table the results of the open loop and autopilot runs are also displayed. These results (columns 2 and 3) illustrate the consistency of the normality conditions imposed in this study on the task numbers.

Design of The Tracking Task - Part II

In this design, the object was to design a different type of target forcing funcion. Figure (3) illustrates the shape of the functions used in this part of the study. In this case the objective was to have forcing functions of varying difficulty. The assumption is that the radius R is a metric of dispersion about the origin defined by equation (1) and tasks with larger R values are more difficult to track. The design of the function in figure (3) is a result of the need to have target tracking tasks that varied the value of R as a function of time. To create the shape of the diagram in figure (3), the following exponential functions were chosen based on a set of time intervals:

Time Interval(seconds)	Function Chosen	
$(t_1, t_2) = (0,5)$	$f_1(t) = \int_0^t a\ e^{-(s-2.5)^2/2\sigma^2} ds$	(2a)
$(t_2, t_3) = (5,10)$	$\dot{f}_2(x) = \dot{f}_1(5) + \int_0^x (-ae^{-(s-7.5)^2/2\sigma^2}) ds$	(2b)
$(t_3, t_4) = (10,15)$	$f_3(y) = f_2(10) + \int_0^y (-ae^{-(z-12.5)^2/2\sigma^2}) dz$	(2c)
$(t_4, t_5) = (15,20)$	$\dot{f}_4(z) = \dot{f}_3(15) + \int_0^z (ae^{-(t-17.5)^2/2\sigma^2} dt)$	(2d)
Where:	$x = t-5$	(3a)
	$y = t-10$	(3b)
	$z = t-15$	(3c)

With some manipulation, the relationships (2a-d) and (3a-c) can be shown to produce the trajectories displayed in figure (3). The value a can be adjusted to sweep out a range of values. Table II illustrates the values chosen for part II of this study:

Table II - Forcing Function Design For Part II

FF #	a	b
1	0.1	.04
2	0.2	.08
3	0.3	.12
4	0.4	.16
5	0.5	.20

Randomization of The Presentation of The Tasks

Reference 6 describes the procedure chosen to ensure that the subjects would not know the order of presentation of the five different tasks at any time during the experiment. This was true for Parts I and II for both the static and stress portions of the experiment.

Apparatus

A 19-foot arm centrifuge (figure 4) was used to establish a 5 Gz stress condition for the subjects. In Air Force applications this acceleration force is in the z direction (down the spine of the subject) and is termed Gz. The centrifuge rotated at an angular speed of 27.5 RPM with the cab vectored at 78 degrees about a line in the z axis of the subject. The subjects wore standard Air Force helmets, gloves, and an Anti-G Suit with a G-valve. The Anti-G Suit-G-valve delivers a specific air pressure to the bladders of the Anti-G Suit.

Training Orientation and Data Exposures

During this training orientation, the subjects were required to asymptote to the five tracking tasks (performance training) and also to acclimate to the G stress (physiological adaptation). In the final design of this experiment, each day's run consisted of five component parts or phases (figure 5 illustrates one day's run for data collection). Phase I comprised of the presentation of the five tracking tasks in the static condition (no stress). Phase II consisted of the presentation of the five tracking tasks at an acceleration stress level of 5 Gz with a 20-second preliminary warm-up run at 4 Gz. After the five exposures at 5 Gz, the centrifuge was brought to a stationary position and the subject again performed five tracking tasks presented in random order in the static condition (Phase III). Phase IV of the daily run consisted of five tracking tasks presented in random order again under the five Gz stress as in Phase II. In Phase V of this experiment, the five tasks were presented in the static mode. Again, as with all the previous tasks, all forcing functions were presented in a random sequence. Four data days were collected after the subject progressed satisfactorily in the indoctrination period. During the data collection phase of the experiment, the subject never experienced more than 300 seconds per day of 5 Gz exposure nor more than two daily exposures per week. After the 4 data days were collected, a questionnaire was administered on the fifth day with the subject sitting in the centrifuge but with no machine motion. The questionnaires recorded subjective ratings of the task difficulty hierarchy.

Questionnaire

One definition of workload (reference 7), indicates that it is a function of increased performance requirements plus additional attention requirements. To get a true subjective evaluation, it was necessary to ask the subjects how they rated the tracking tasks. On the last day of the experiment the subjects were presented 25 tasks in random order. After the first task, each subject was asked to compare the task he was presently tracking with the previous one. The subject was asked whether the present task was more difficult, less difficult, the same, or not possible to rate. Thus the subject, was not knowledgeable as to which forcing function number was presented and would only give relative ratings between tasks.

RESULTS

CDF Performance Results From The Data

As discussed previously, after the 5-day indoctrination period the seven subjects tracked until they trained to an asymptotic level of performance for the different tracking tasks. One criterion used to define asymptotic performance is that on three consecutive days, the RMS performance scores do not decrease more than 5% on daily exposures of 25-50 static presentations of the five random targets per day. After this level was reached, the subjects were assumed to be trained. In part I of this study, the first question to be asked concerns the adaptation of the subjects to learning the tracking task and accclimation to G levels.

To address the question of learning and adaptation to stress, a table based on error scores was constructed across all seven subjects and four replications of each stress condition. Table III illustrates these results:

Table III - Stress Data $\overline{x}/\sigma$ Ratios For 7 Subjects, 2 Replications/Day

	Day 1	Day 2	Day 3	Day 4
ff#1	1.9	5.2	2.9	4.6
ff#2	2.6	4.9	3.6	4.0
ff#3	5.3	4.5	13.1	10.6
ff#4	8.2	10.0	17.1	13.5
ff#5	10.1	5.9	24.3	20.4

If any trends did exist in the data runs, due either to further performance training (reduction of tracking error) or possibly to further acclimation to G stress (physiological adaptation), they would be shown by a gradual increase in the ratio $\overline{x}/\sigma$ across a row for a given forcing function number. Since there appear to be no apparent trends for this stress acclimation, it is assumed that the subjects had adjusted to a steady state physiological conditioning and tracking performance level.

The next question to be addressed here is whether the forcing function number was correlated with measures of performance degradation. From the CDF figures (similar to figure 6), using data from all seven subjects (five replications), it was desired to conduct tests to investigate if a significant performance decrement exists dependent on forcing function number for both the static or the stress conditions. The following statistical test would determine this effect:

$$H_0: \quad \mu_{i+1} > \mu_i \qquad i=1,\ldots 4 \qquad (4a)$$

versus

$$H_1: \quad \mu_{i+1} \leq \mu_i \qquad i=1,\ldots 4 \qquad (4b)$$

where μ corresponds to the 0.5 line on the CDF in figure 6. This figure is illustrated here to show how the median point μ is obtained. This corresponds to a "median window" size for the tracking error signal. The test was conducted for both the static data and the stress data. The results using a t statistic are displayed in Table IV:

Table IV

Hypothesis test on values	$\overline{t}$ for static data	p	$\overline{t}$ for stress data	p
$ff_2 > ff_1$	5.174	<.01	11.48	<.01
$ff_3 > ff_2$	18.39	<.01	19.91	<.01
$ff_4 > ff_3$	7.15	<.01	9.64	<.01
$ff_5 > ff_4$	4.55	<.01	5.04	<.01

The $\overline{t}$ values used in Table IV were the t statistic (2-tailed test) for correlated data (references 8 and 9) which satisfies:

$$\bar{t} = Md \Big/ \sqrt{\frac{\sum_{i=1}^{n} x^2{}_d}{n(n-1)}} \tag{5}$$

where Md = the mean of the n differences of paired observations and x_d = the deviation of a difference from the mean of the differences. This test is for paired samples; they are not independent but are correlated due to the five replications involved with all seven subjects in the static case, and four replications with the stress data. The results from Table IV indicate performance decrements correlated with the forcing function number. The performance decrement is significant at an .01 level for an increase of forcing function number in both the static and stressed condition. The test given here corresponds to changes in medians (i.e., the 0.5 point on the CDF curve). Using the CDF method, this analysis could have been performed for any window size or any other point on the CDF curve. This is emphasized here because in other types of applications it may be desirable to look at a specified level of the CDF function (e.g. $CDF \neq 0.5$) or at a specified window size. Finally, the tests illustrated here hold over both the static and stress conditions.

Another question to be addressed is whether the effects of the physiological stress induce a performance change for each task number. Using the data from the seven subjects and four replications of the stress condition, Table V illustrates the effects of stress on tracking performance.

Table V Comparisons of Stress vs Static Conditions

Hypothesis test on μ values	$\bar{t}$ for this test	p
ff_1 stress > ff_1 static	3.34	$<.05$
ff_2 stress > ff_2 static	1.54	$<.10$
ff_3 stress > ff_3 static	2.81	$<.05$
ff_4 stress > ff_4 static	5.83	$<.01$
ff_5 stress > ff_5 static	3.14	$<.05$

The t statistic used in this test is the same as in equation 5. One can now see the impact on performance degradation as noted by the effect of stress on tracking in Table V.

In part II of this study, it was desired to study this sensitivity effect for the second class of tracking tasks. Table VI illustrates the actual error score results for Part II as well as the equivalent values found in Part I:

Table VI

	Part I				Part II			
	Static		Stress		Static		Stress	
ff#	$\bar{x}$ mean	$\bar{x}$ s.d.	$\bar{x}$ mean	$\bar{x}$ s.d	$\bar{x}$ mean	$\bar{x}$ s.d.	$\bar{x}$ mean	$\bar{x}$ s.d.
1	94.8	26.4	127.7	29.6	22.67	3.87	37.73	7.14
2	162.4	48.7	179.0	33.2	38.42	6.71	59.14	12.02
3	283.9	49.4	306.3	29.7	54.32	11.40	79.91	14.88
4	388.6	22.9	414.7	37.4	106.32	26.63	121.77	34.89
5	624.5	23.9	649.5	26.4	223.09	38.82	239.96	48.58

The results of the statistical tests indicate a performance decrement under various conditions. The subjective data from the questionnaire are presented next.

Results From The Questionnaire

In the questionnaire, the subjects were asked to compare the relative difficulty of the task they were presently tracking with the previous task. Responses of "more difficult", "less difficult", "the same", or "couldn't tell" were then correlated with the task numbers presented. These results are displayed in table VII for Parts I and II of this study:

Table VII - Correlation of Task Numbers with Subjective Responses*

	Part I		Part II	
Subject #	# Correct/Total	% Correct	# Correct/Total	% Correct
1	24/25	96%	21/25	84%
2	23/25	92%	25/25	100%
3	24/25	96%	19/25	76%
4	25/25	100%	25/25	100%
5	23/25	92%	23/25	92%
6	23/25	92%	24/25	96%
7	23/25	92%	25/25	100%

The subjects also commmented that as they were presented tasks with higher forcing function numbers, the tracking tasks required more attention. This corresponds to the description of workload cited earlier in which higher performance requirements coupled with more stringent attention requirements increase workload.

SUMMARY AND CONCLUSIONS

This study used sensitive tracking tasks to evaluate performance degradation under acceleration stress. The tasks designed here had to satisfy certain criteria. First they had to be zero mean, constant variance, sum of sines. Second, open loop scores for all five tasks had to be identical. Third, the autopilot runs also had to yield a consistent score for all five tasks. When the human was tracking these tasks, however, a performance decrement had to be observed dependent on forcing function number for static tracking. In addition, the performance decrement had to occur as a function of the experimental conditions stress versus non-stress for each forcing function number.

At the conclusion of the experiment the subjects were given a questionnaire to rate the different tasks. Subjective ratings of each task in order of difficulty were necessary in order to verify the workload definition used here, which requires both a performance decrement and an attention requirement for arranging tracking tasks in order of increasing difficulty.

---

* Due to different subject pools in Parts I and II of this experiment, subject #N in Part I may not be the same person as Subject #N in Part II (N=1,.,7).

REFERENCES

1. Cooper, G. E.,1957, "Understanding and Interpreting Pilot Opinion", Aeronautical Engineering Review, March, pp. 47-56.

2. Cooper, G.E. and Harper, R.P., Jr., "The Use Of Pilot Rating In The Evaluation of Aircraft Handling Qualities" AGARD Report 567, April, 1969.

3. Repperger, D.W., Ward, S.L., Hartzell, E.J., Glass, B.C., and Summers, W.C., 1979,"An Algorithm To Ascertain Critical Regions of Human Tracking Ability", IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-9,No. 4,April, pp.183-196.

4. Jex, H.R., McDonnell, J.E., and Phatak, A.V., "A Critical Tracking Task For Man-Machine Related To The Operator Effective Delay Time I: Theory and Experiments With A First Order Divergent Controlled Element", NASA CR-616, October, 1966.

5. Verplank, W.L., 1977, "The Facilitating Effects of Uncertainty In Long-Term Manual Control", Proceedings of The Thirteenth Annual Conference on Manual Control, June 15-17, MIT, pp. 101-105.

6. Repperger, D.W., Rogers,D.B., "A Task Difficulty - G Stress Experiment", Submitted for Publication.

7. Levison, W.H., 1979, Comments on Workload, at the Fifteenth Annual Conference on Manual Control, Wright State University, Dayton, Ohio.

8. Snedecor, G.W. and Cockran, W.G., 1967, "Statistical Methods", The Iowa State University Press, 6th. Edition.

9. Guilford, J.P., 1956, "Fundamental Statistics in Psychology and Education", 3rd. Edition, McGraw-Hill.

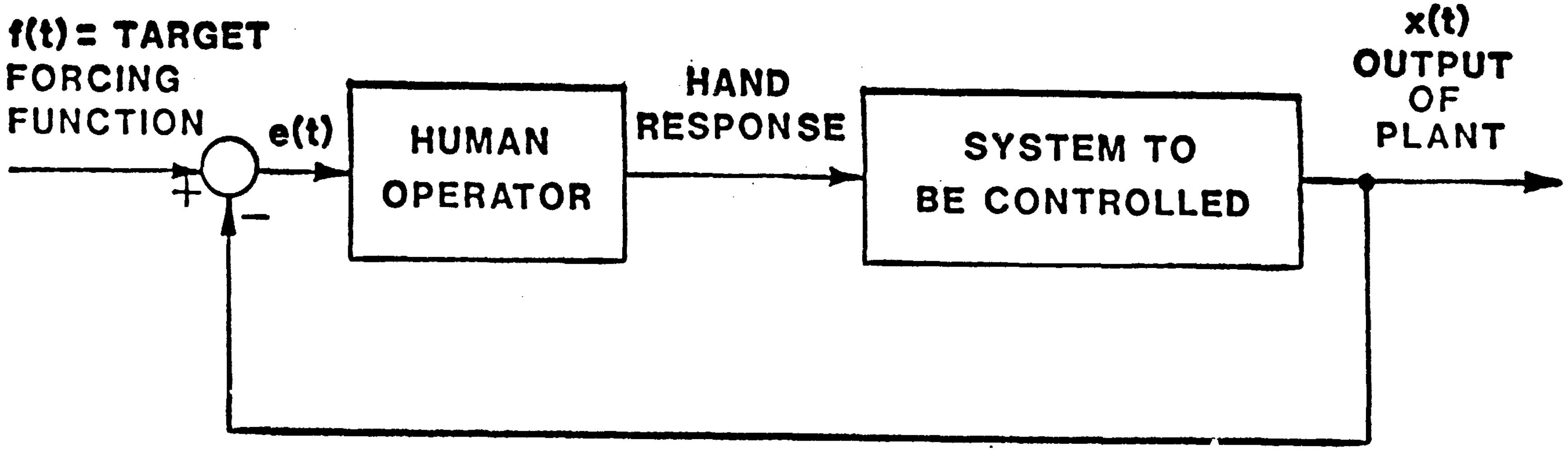


Figure (1) - THE MAN-MACHINE SYSTEM

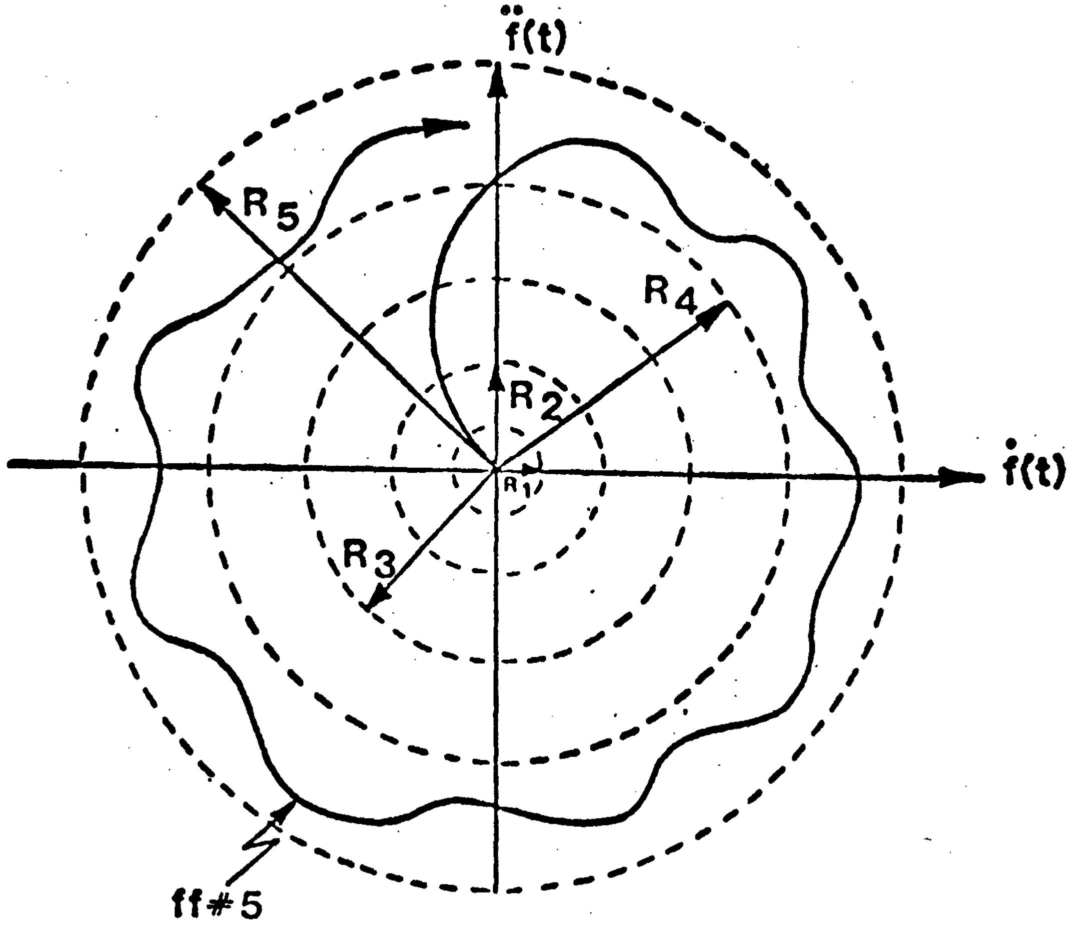


Figure (2) - A METRIC OF TRACKING DIFFICULTY

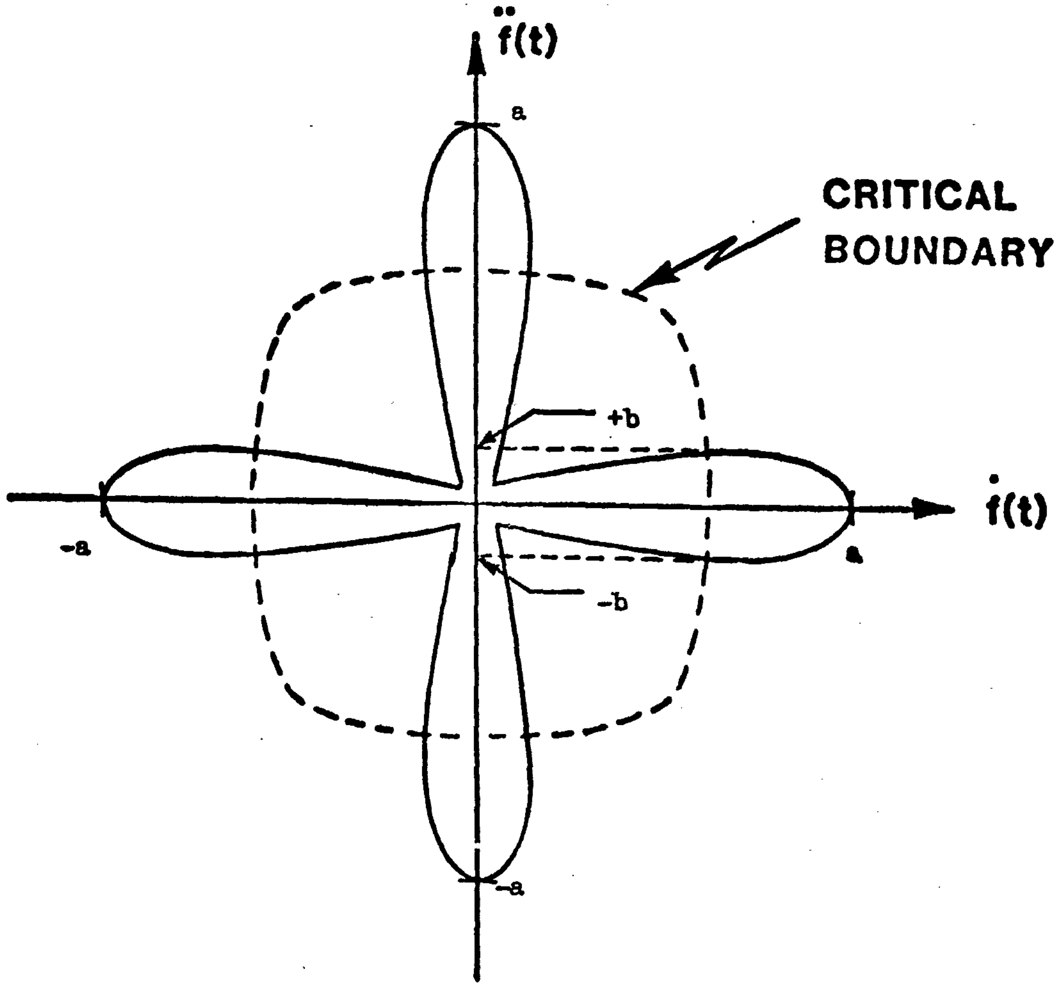


Figure (3) - SWEEP FUNCTIONS TO INVESTIGATE CRITICAL BOUNDARIES

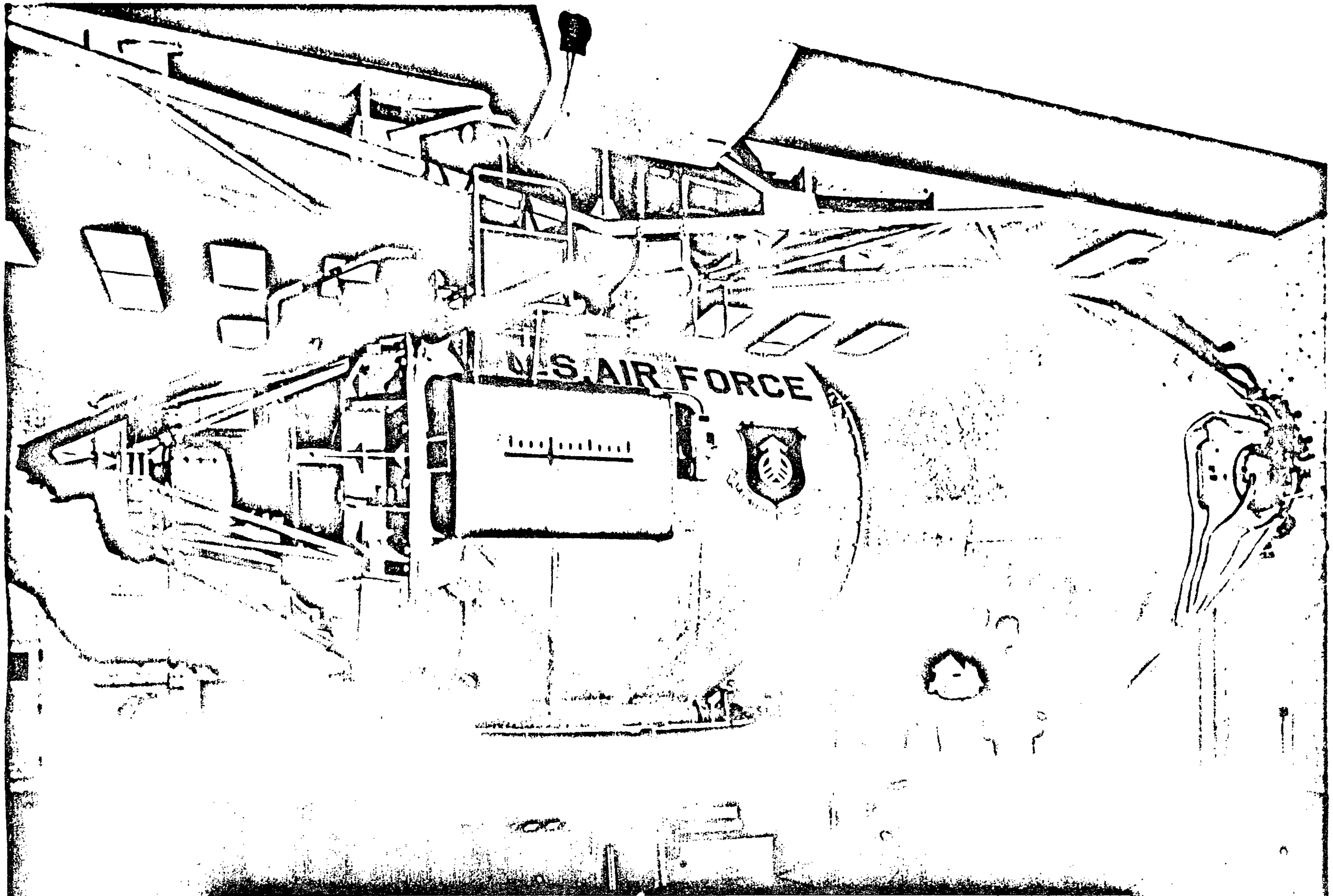


Figure (4)

THE DYNAMIC ENVIRONMENTAL SIMULATOR (D.E.S.) CENTRIFUGE
THE AIR FORCE AEROSPACE MEDICAL RESEARCH LABORATORY
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

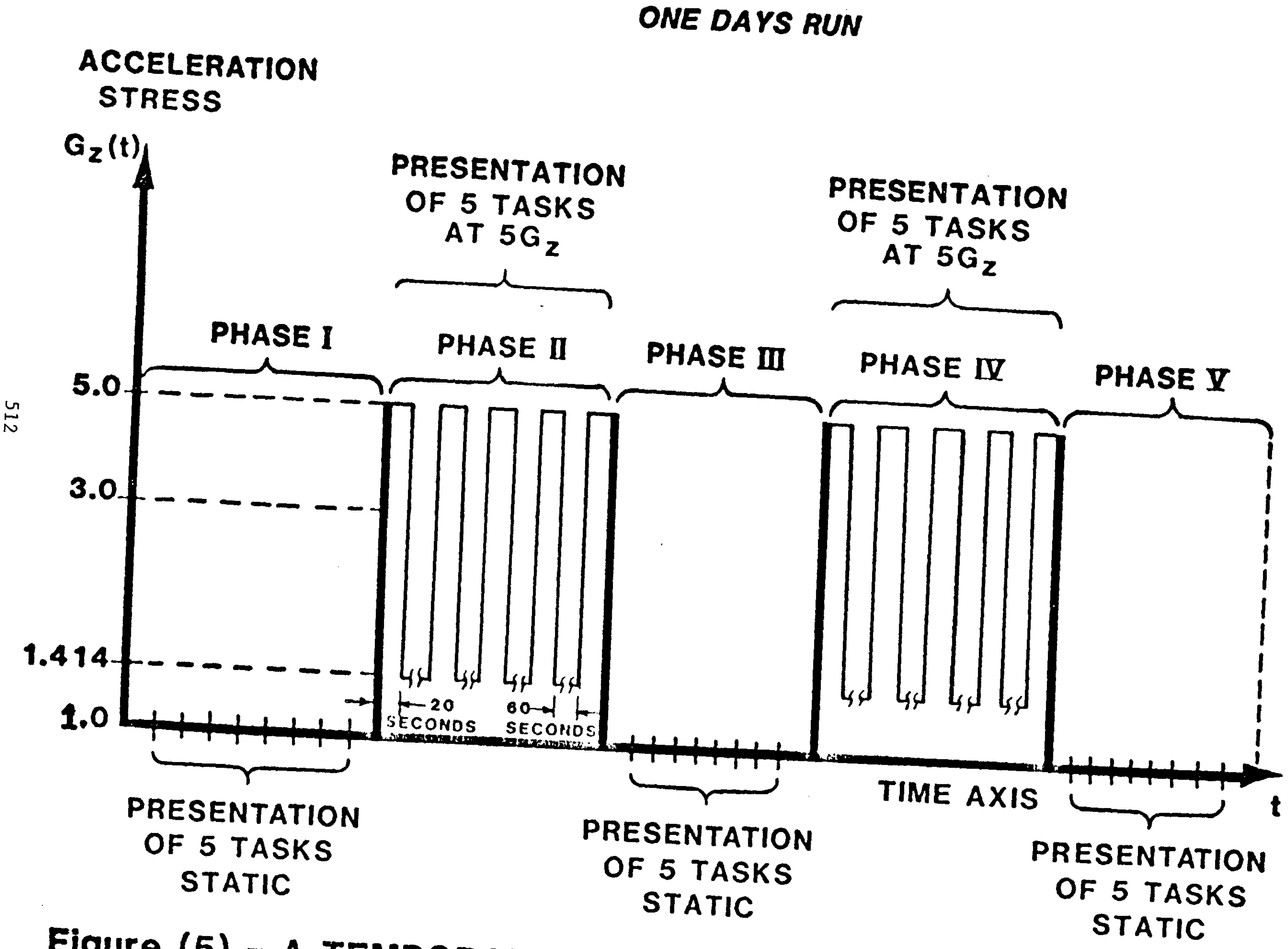


Figure (5) - A TEMPORAL DESCRIPTION OF THE G STRESS EXPOSURE

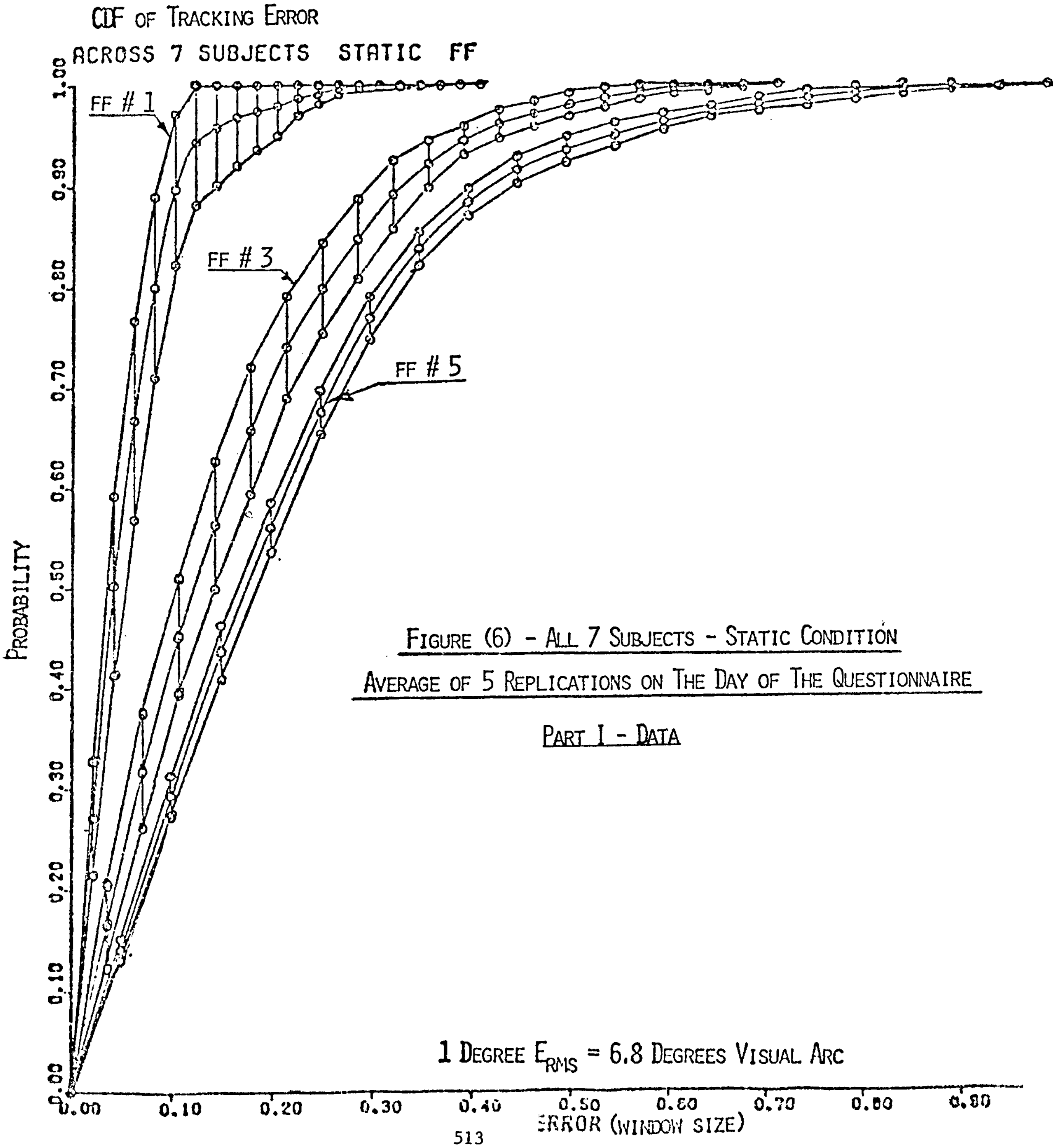


Figure (6) - All 7 Subjects - Static Condition

Average of 5 Replications on The Day of The Questionnaire

Part I - Data

PROJECT: HEL Cockpit Lighting Compatibility (HELCLIC)

PRINCIPAL INVESTIGATOR: MAJ Thomas L. Frezell

ABSTRACT

1. Background. The use of night vision goggles is considered imperative to developing a near term night operational capability for US Army aviation assets. Cockpit lighting systems in current US Army fleet helicopters interact adversely with night vision goggles.

The US Army Human Engineering Laboratory (HEL) has developed several modifications to cockpit lighting systems which are intended to alleviate the present goggle/lighting incompatabilities. These systems have been tested under laboratory conditions. The JUH-1H instrumented helicopter has been modified with two of the most promising candidates.

2. Objective. Determine the effects of conventional versus modified cockpit lighting on pilot performance while flying with night vision goggles. (AN/AVS-6)

3. Participants. The participants in this experiment will be rated aviators from both the HEL and Phillips Army Airfield. The N will consist of from five to eight pilots.

4. Apparatus. The apparatus will consist of:

a. JUH-1H Instrumented Helicopter
b. Night Vision Goggles
c. The current cockpit lighting and modified cockpit lighting systems.

5. Procedures and Methodology. Pilot subjects will be briefed on the general nature of the project and the flight maneuver to be performed. The flight maneuvers will consist of a 360° right hovering turn at two altitudes, 3 ft and 50 ft. These maneuvers will be flown for all test conditions (i.e., three cockpit lighting conditions with and without goggles.) Initially, each subject will complete the flight maneuvers under the three lighting conditions without goggles to preserve their dark adaptation and conserve flight time. Dark adaptation will be accomplished by having each subject dark adapt with red goggles for thirty (30) minutes prior to initiation of flight tests. The order of presentation of cockpit lighting will be randomized across subjects to preclude order effects.

Each subject will be asked to bring the aircraft to the appropriate altitude (3 ft and 50 ft) within ±1 ft (Aircrew Training Manual Standard) and complete a 360° right hovering turn about

the mast of the helicopter maintaining altitude within ± ft, drift not to exceed 2 ft from pivot point and a constant rate of turn, not to exceed 90° in 4 seconds.

The data to be recorded are absolute altitude (radar) Z; Doppler Position Data (x & y), Flight Control Inputs (cyclic, pedals and collective). Derived measure will be rate of turn completed.

All maneuvers to be performed are standard flight maneuvers. The 3 ft hover is accomplished during every flight; while the 10 ft and 50 ft maneuvers are accomplished prior to performing OGE (out of ground effect) flight. Each subject will provide three sets of maneuvers for each altitude.

6. Experimental Design. The experiment will be conducted as a 3x2x2 factorial design with repeated measures. The independent variables will be (1) Flight Maneuvers; (2) Goggles/No Goggles and (3) Cockpit Lighting Systems. The dependent variables will be (1) Hover Position; (2) Turn rate; (3) Flight Control Activity and (4) Absolute Altitude Variations. Each subject will have three trials per test condition.

The order of presentation of lighting conditions will be randomized across subjects.

7. Data Analysis. The flight data is recorded on magnetic tape and consists of: the altitude above ground, the position over ground, aircraft heading and flight control activity. The control activity consists of the cyclic latitude, the cyclic longitude, the collective and the pedal movements. The appropriate sensors are sampled 20 times per second and are recorded on the analog tape by the helicopter instrumentation package (reference, TM 17-80). The recorded data is transferred to digital tape after the flight and read into a data file on the Cyber 7600 computing facility here at Aberdeen Proving Ground, Maryland.

Event markers will be set during the test to mark the start and end of each test run for the conditions of a subject at a given test altitude, cockpit lighting, goggles usage, and replication. The setting of these toggle switches will be recorded on the flight data tape and, in turn, read into the Cyber data file.

The mean and standard deviation shall be computed from the Cyber data file for each flight activity of each test run. The means will be used as cell entries in a 3x2x2 random block Analysis of Variance with repeated measure for each flight activity. The analysis shall be for the independent variable of altitude, cockpit lighting and goggle usage.

SYSTEMS TECHNOLOGY, INC.

2672 BAYSHORE FRONTAGE ROAD • MOUNTAIN VIEW, CALIFORNIA 94043 • PHONE [415] 961-4674

In reply refer to:

Paper No. 311

OVERVIEW OF WORK IN PROGRESS ON NON-INTRUSIVE ASSESSMENT OF PILOT WORKLOAD AND PILOT DYNAMICS

Robert K. Heffley
Warren F. Clement
Wayne F. Jewell

January 1982

For Presentation at the AIAA Workshop on
Flight Testing to Identify Pilot Workload and Pilot Dynamics

January 19-21, 1982

Air Force Flight Test Center
Edwards Air Force Base, California

OVERVIEW OF WORK IN PROGRESS ON NON-INTRUSIVE ASSESSMENT OF PILOT WORKLOAD AND PILOT DYNAMICS

INTRODUCTION

The purpose of this paper is to describe one general approach to non-intrusive assessment of pilot workload and pilot dynamics, how this assessment is carried out, and recent and on-going projects which involve its application. We begin with a general closed-loop formulation of piloting technique which can include both psychomotor and cognitive aspects. Examples are given showing how this basic closed-loop structure can be applied to a variety of flight tasks with the help of several quantitative examples. We then go on to consider the major objectives of work in progress, some of the tools and concepts used in non-intrusive piloting technique assessment, and a number of the important lessons learned in applying various procedures. Finally, we summarize some of the recent and on-going projects involving non-intrusive piloting technique assessment procedures.

BACKGROUND

Historically we have depended heavily upon the pilot to provide us with direct and explicit information concerning manual control strategy, perceptual dynamics, and pilot workload. The last item is often expressed in terms of a pilot opinion rating. Our dependence upon pilots to assess their own actions, perceptions, and degree of stress, however, can interfere with what it is that we are trying to measure. For example, assessment of workload using a Cooper-Harper rating first requires a pilot with special training in using the rating scale. This evaluation pilot may not have sufficient background and skills for the particular mission or aircraft type being evaluated. We have found that pilot control strategy is even more difficult to determine from direct interrogation of the pilot. While it is relatively easy for most pilots to relate to the notion of a basic closed-loop structure representing piloting technique in performing a specific task, it is difficult for them to quantify such a loop structure. The same observations are true for perceptual aspects of the pilot dynamics. Although the pilot can identify aircraft states used in accomplishing a particular task, it may not be possible to describe geometric cue patterns or sensations actually used or their relative amount of use.

It is therefore useful to search for ways of measuring and quantifying pilot dynamics and pilot workload connected with the execution of given tasks which do not depend on direct interrogation of the pilot but rather upon how the pilot actually performs or executes tasks in a flight environment.

AN OVERVIEW OF NON-INTRUSIVE PILOT ASSESSMENT

Figure 1 shows a general overview of manual task execution in terms of the pilot-vehicle system and how this system can be measured in order to obtain a non-intrusive assessment of pilot dynamics and pilot workload. Task execution is expressed in terms of a general feedback control system. A wide spectrum of piloting tasks and styles of task execution can be cast in terms of this same general topology. In addition to familiar continuous psychomotor tasks (e.g., tracking a gunsight pipper error, a localizer, or a glide slope) we can also address discrete maneuvers (e.g., landing flare; airspeed, altitude, or heading change; and responding to traffic control commands); discrete tasks (e.g., configuration selection, radio tuning, and mode switching); communications within or without the cockpit; and, finally, checklist execution procedures.

Note in Fig. 1 that the interface between the pilot and the vehicle is between the elements of control strategy and the vehicle dynamics. Perceptual dynamics are often regarded as part of the pilot dynamics but could be alternatively considered as part of the vehicle dynamics under

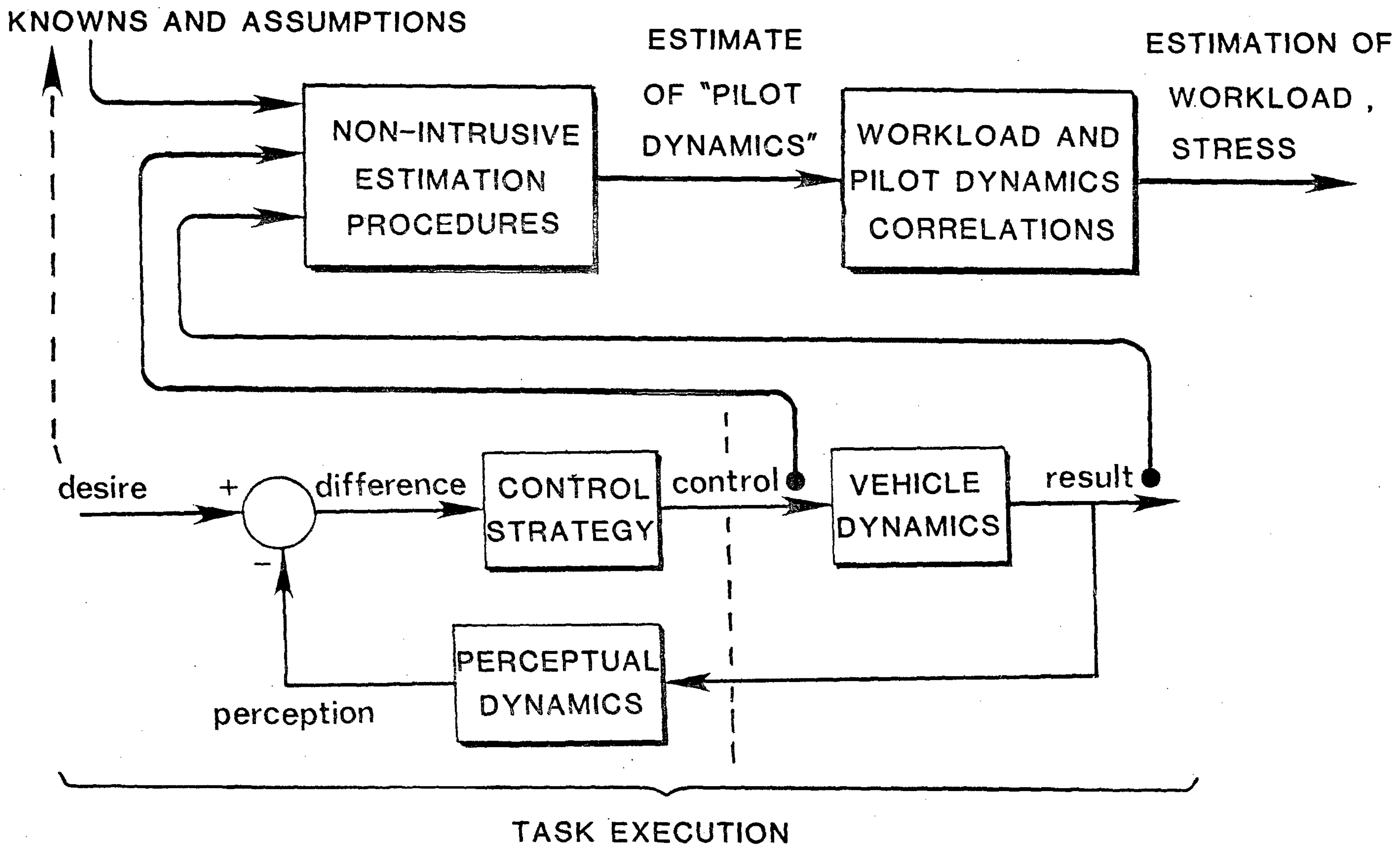


Figure 1. Non-Intrusive Pilot Assessment

certain conditions, such as inclusion of simulator motion washouts or visual system properties.

Task execution in general consists of transforming some pilot desire or pilot command to a result which is often expressible in terms of a vehicle state. Depending upon the difference between what a pilot desires and the pilot's perception of what results, a control strategy is called upon to define a control action. The control in turn, acting through the vehicle dynamics, yields the result of the task execution. The loop is finally completed with the comparison of desire with result via the perceptual dynamics.

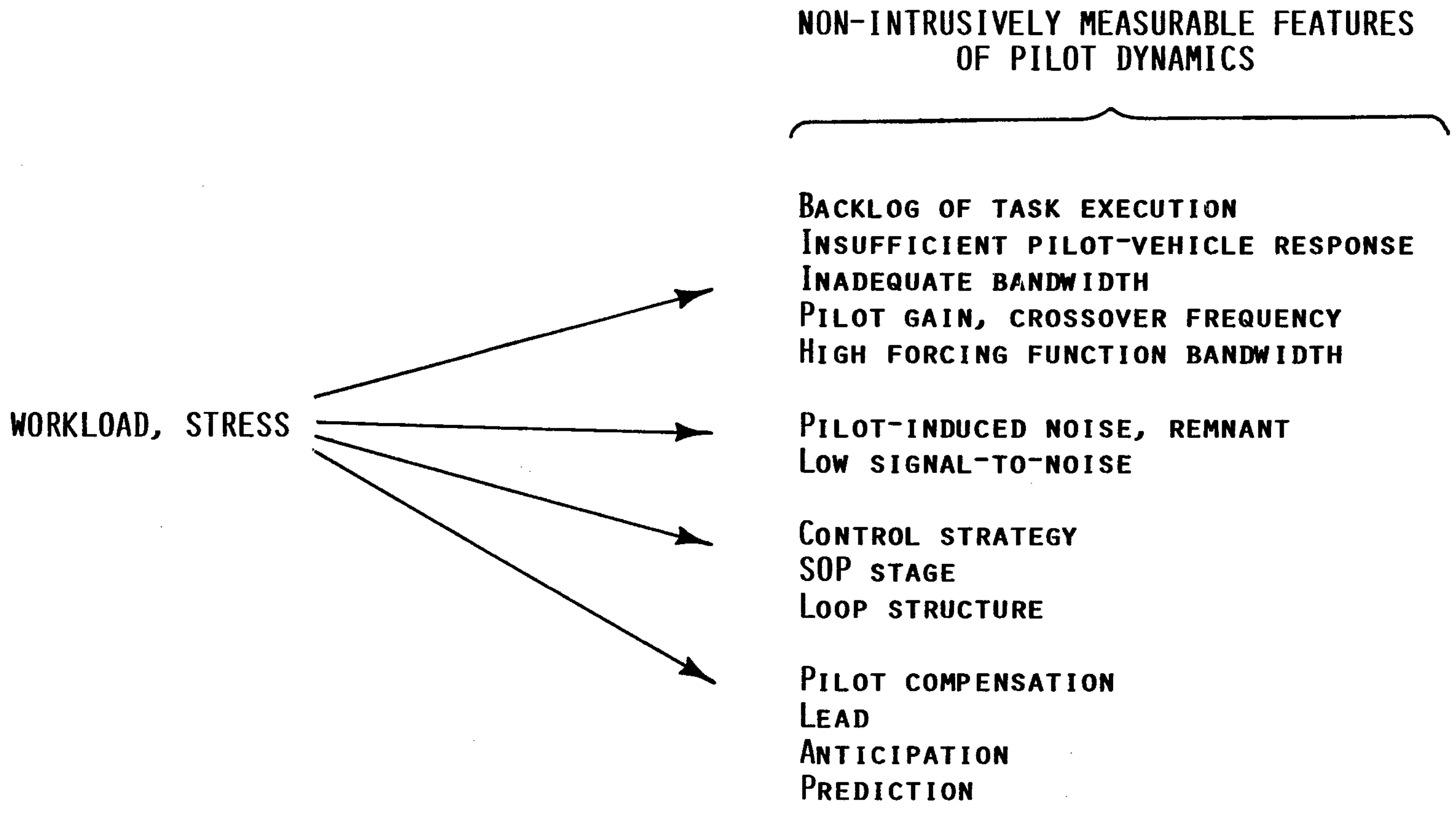
In general, non-intrusive estimation procedures make use of the pilot's control and task result, both of which are usually easily quantified and measurable. The control and result must, however, be combined with knowns and assumptions concerning the various elements of task execution; i.e., we might choose to assume a form of control strategy and judge that we know the vehicle dynamics and approximate perceptual dynamics. Skillful utilization of such information often permits us to make a good estimate of pilot dynamics. Thus, as a result of observing how a pilot is executing a task, we can frequently estimate the control strategy being used and the perceptual dynamics involved in the task.

It should be noted that we could also identify simultaneously the vehicle dynamics. It is sometimes desirable to do so, and we will mention how we can separately measure vehicle dynamics and pilot dynamics while using the same measurement data. The final step is to take the estimated pilot dynamics and, using known correlations between workload and pilot dynamics such as listed in Table 1, estimate the pilot workload or stress. Thus in this indirect set of procedures we have avoided any direct interrogation of the pilot or hindrance of piloting activities.

Figures 2 through 6 illustrate a number of examples of this task execution topology for which some quantification has been derived from simulator, flight, and analysis. The next section illustrates some of the varieties of tasks which can be expressed in this general form; however, pilot-vehicle elements are assigned other more appropriate names.

TABLE 1

CORRELATES OF WORKLOAD, STRESS



WORKLOAD, STRESS

NON-INTRUSIVELY MEASURABLE FEATURES
OF PILOT DYNAMICS

BACKLOG OF TASK EXECUTION
INSUFFICIENT PILOT-VEHICLE RESPONSE
INADEQUATE BANDWIDTH
PILOT GAIN, CROSSOVER FREQUENCY
HIGH FORCING FUNCTION BANDWIDTH

PILOT-INDUCED NOISE, REMNANT
LOW SIGNAL-TO-NOISE

CONTROL STRATEGY
SOP STAGE
LOOP STRUCTURE

PILOT COMPENSATION
LEAD
ANTICIPATION
PREDICTION

— CAN APPROACH PILOT WORKLOAD VIA MEASUREMENT OF PILOT DYNAMICS!

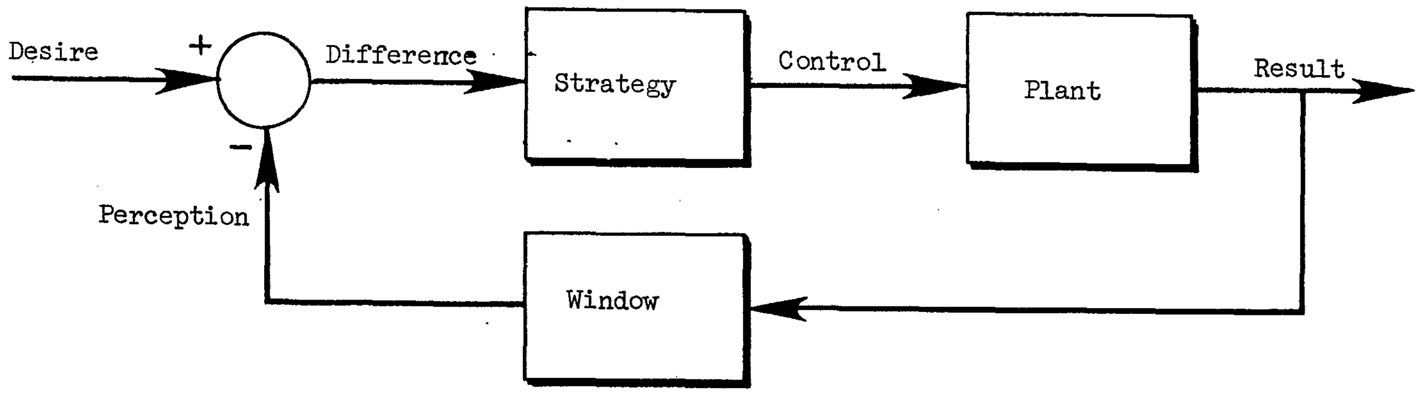


Figure 2. Organization of General Task Execution

DYNAMICS: RESPONSE TIME AND SETTLING

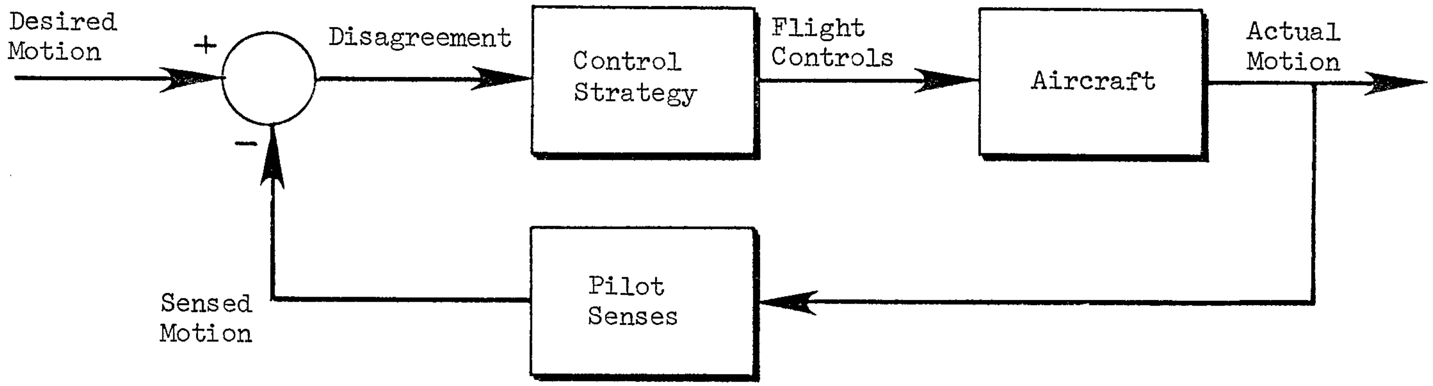


Figure 3. Flight Maneuver Execution

DYNAMICS: TIME AND CERTAINTY OF INTERPRETATION

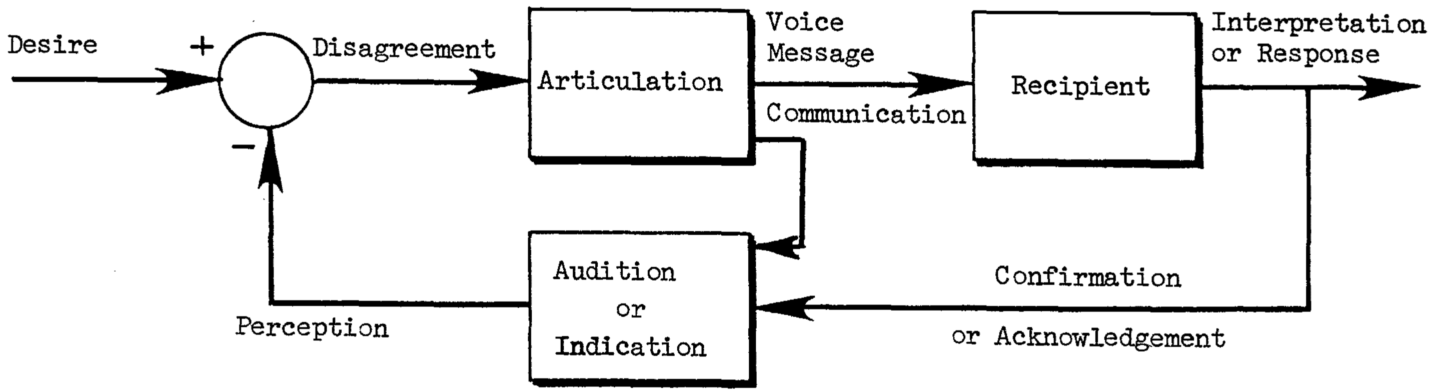


Figure 4. Communication

DYNAMICS: TIME AND SURENESS OF EXECUTION

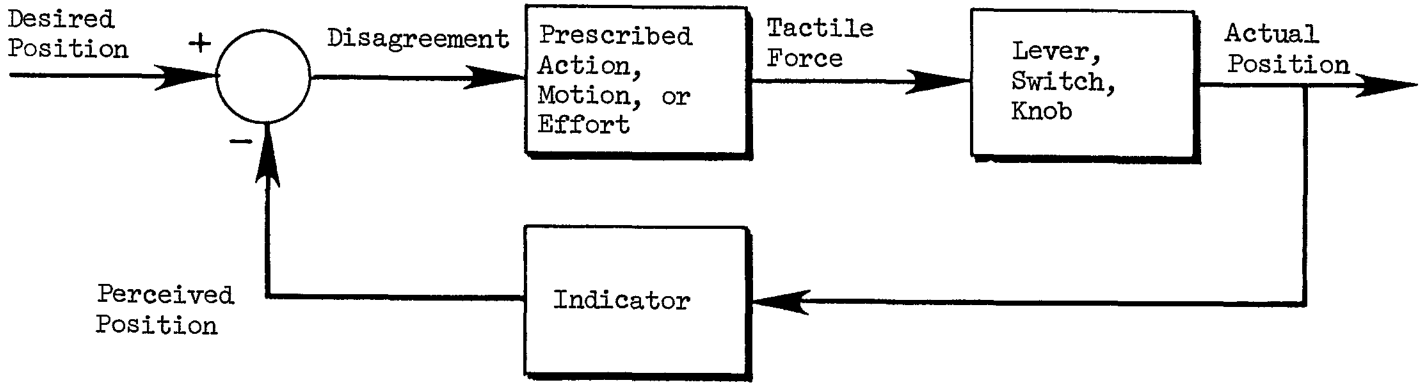


Figure 5. Throwing Switches, Selecting Configuration Levers, and Tuning Radios

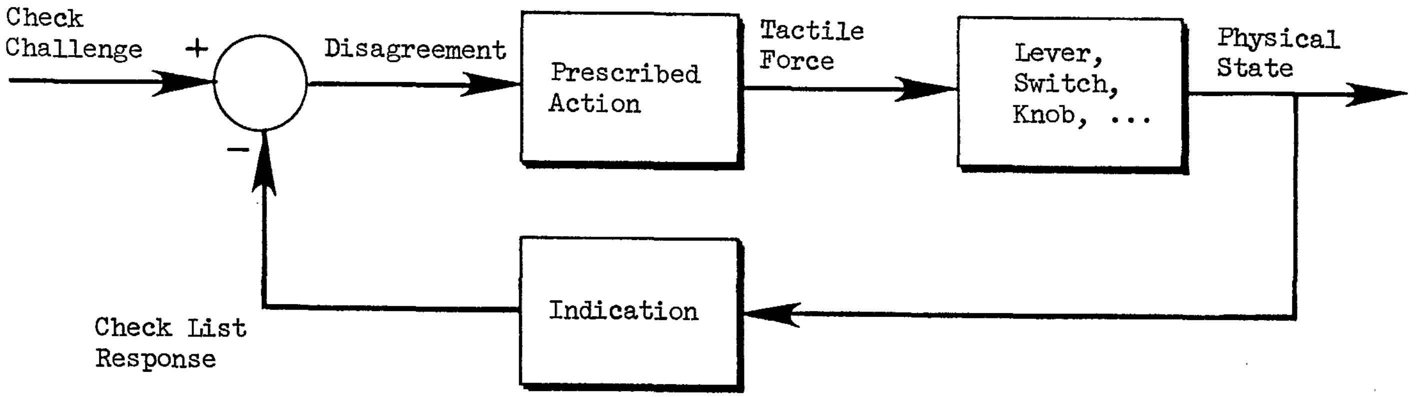


Figure 6. Execution of Checklist Item

SPECIFIC TASK ANALYSES

Few examples exist of even partially quantified pilot-vehicle loop structures for full-task analyses. A review of the literature shows an emphasis on loop structures for partial task analysis, and this is limited mainly to inner-loop aspects. Figures 7 through 13 (excerpted from Ref. 1) depict some of those few cases where relatively complete outer loop tasks have been considered and, to some degree, quantified. Note that each task is described first verbally and then in terms of a feedback control block diagram. In each case an attempt is made to focus only on the primary task loop. In the interest of simplicity, supporting or secondary loops are omitted (e.g., for a speed change maneuver only the longitudinal control axis is addressed —— supporting pitch attitude control and control of other axes are not shown explicitly.) In each example the input to the task execution block diagram is the pilot's desired state, and the output is the actual resulting state. In all cases there is an explicit control strategy shown, but an explicit perceptual block is included only for the helicopter decelerating-approach-to-hover (Fig. 12). Where no perceptual block is shown, perfect pilot perception of a state is assumed.

OBJECTIVES OF NON-INTRUSIVE PILOT ASSESSMENT

The objectives of the work in progress are, first, to assess in quantitative terms the control strategy and the perceptual dynamics for mission-oriented flight tasks without employing measurement procedures which hinder the pilot; i.e., we are looking for non-intrusive measurement techniques. Furthermore, we are trying to develop and refine non-intrusive piloting technique measurement procedures for routine flight and simulator use. The utility of such measurements is much less if one cannot use them routinely with ease; and this implies that one has to look for simple, fast computational procedures. The next objective is to clarify and quantify already established relationships between non-intrusively measured pilot-vehicle dynamics and pilot workload, examples of which

Verbal Description

Adjust bank angle by applying appropriate pressure to the lateral cyclic stick. Monitor roll response using either the actual horizon or the cockpit artificial horizon.

Feedback Control Description

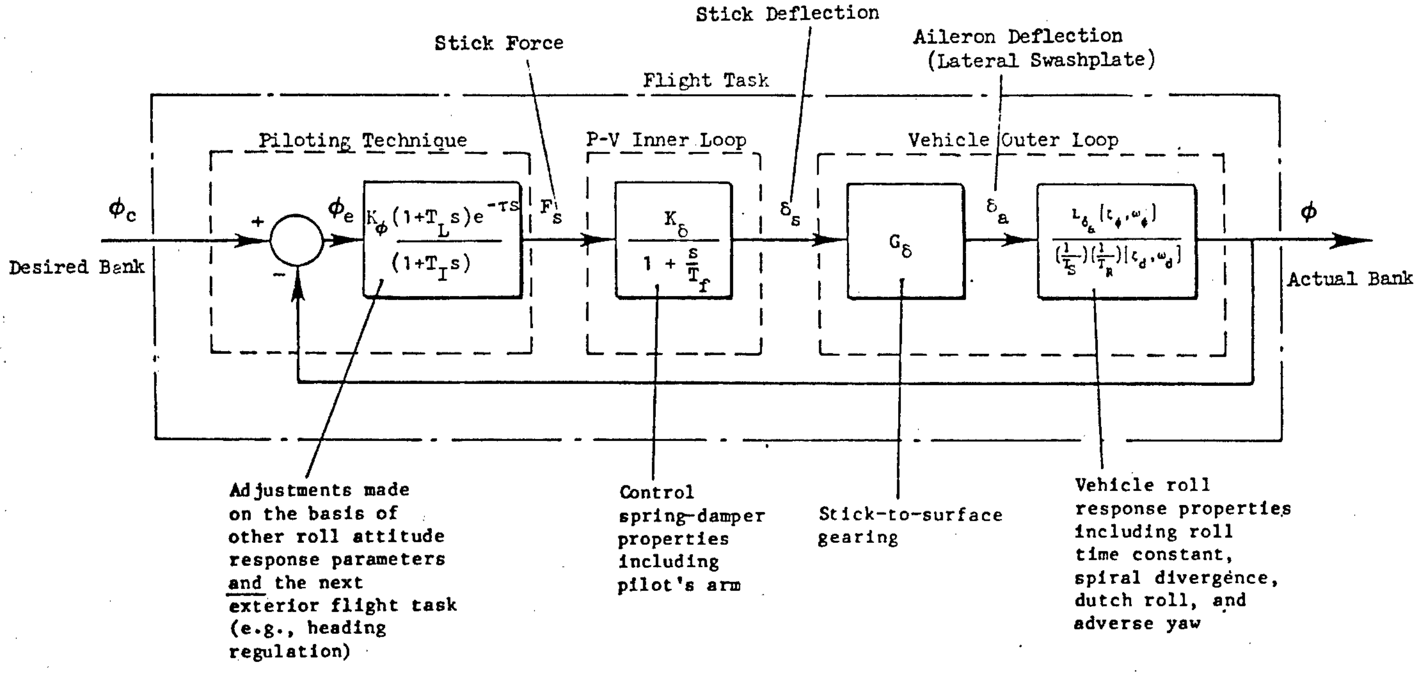


Figure 7. Bank Angle Regulation Task and Piloting Technique

Verbal Description (Ref. 2)

When a deviation from the desired heading occurs, refer to the attitude indicator and smoothly establish a definite angle of bank which will produce a suitable rate of return. As a guide, the bank attitude change on the attitude indicator should equal the heading deviation in degrees not to exceed 30 deg. (For example, if the heading deviation is 10 deg, then 10 deg of bank would produce a suitable rate of correction.) This guide is particularly helpful during instrument approaches at relatively slow airspeeds. At higher true airspeeds, a larger angle of bank may be required to prevent a prolonged correction. A correction to a heading deviation of 2 deg to 5 deg may be accomplished by application of rudder.

Feedback Control Description

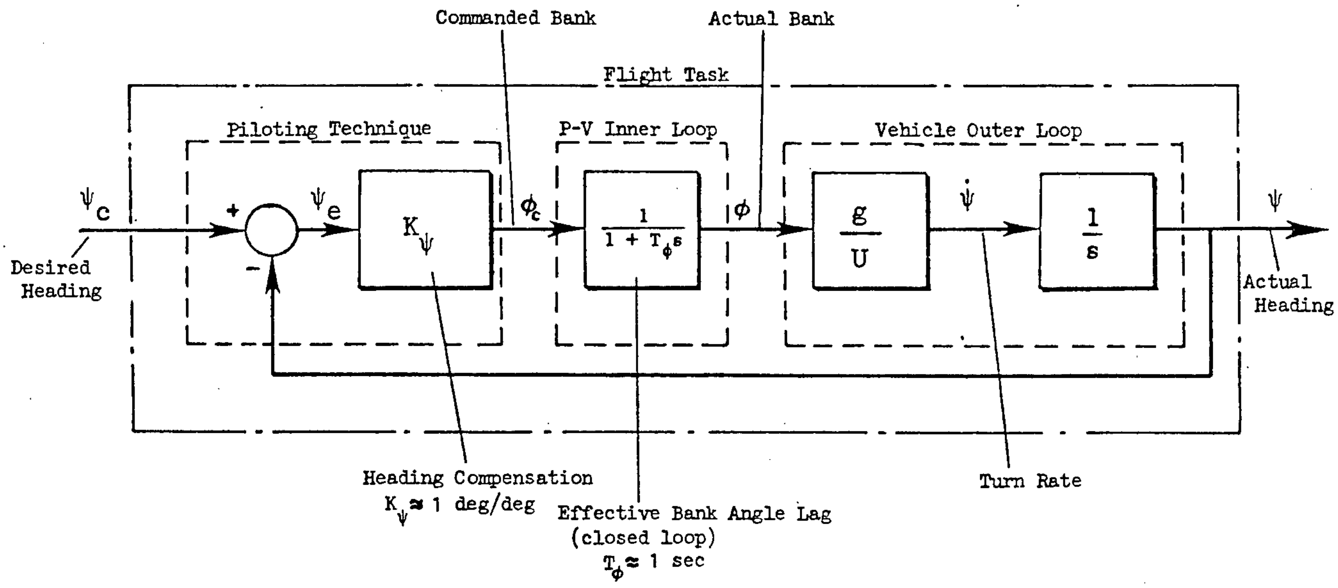


Figure 8. Heading Regulation Task and Piloting Technique

Verbal Description (Ref. 2)

To enter a turn, you should refer to the attitude indicator while applying smooth and coordinated control pressures to establish the desired angle of bank. Bank control should then be maintained throughout the turn by reference to the attitude indicator. Cross-check the heading indicator or turn needle to determine if the angle of bank is satisfactory. Trim may be helpful during prolonged turns to assist in aircraft control.

To roll out of a turn on a desired heading, a lead point must be used. The amount of lead required depends upon the amount of bank used for the turn, the rate the aircraft is turning, and your rollout rate. As a guide, a lead point of approximately 1/3 the angle of bank may be used. With experience and practice a consistent rate of rollout can be developed. A lead point can then be accurately estimated for any combination of angle of bank and rate of turn. Make a note of the rate of movement of the heading indicator during the turn. Estimate the lead required by comparing this rate of movement with the angle of bank and the rate of rollout.

Feedback Control Description

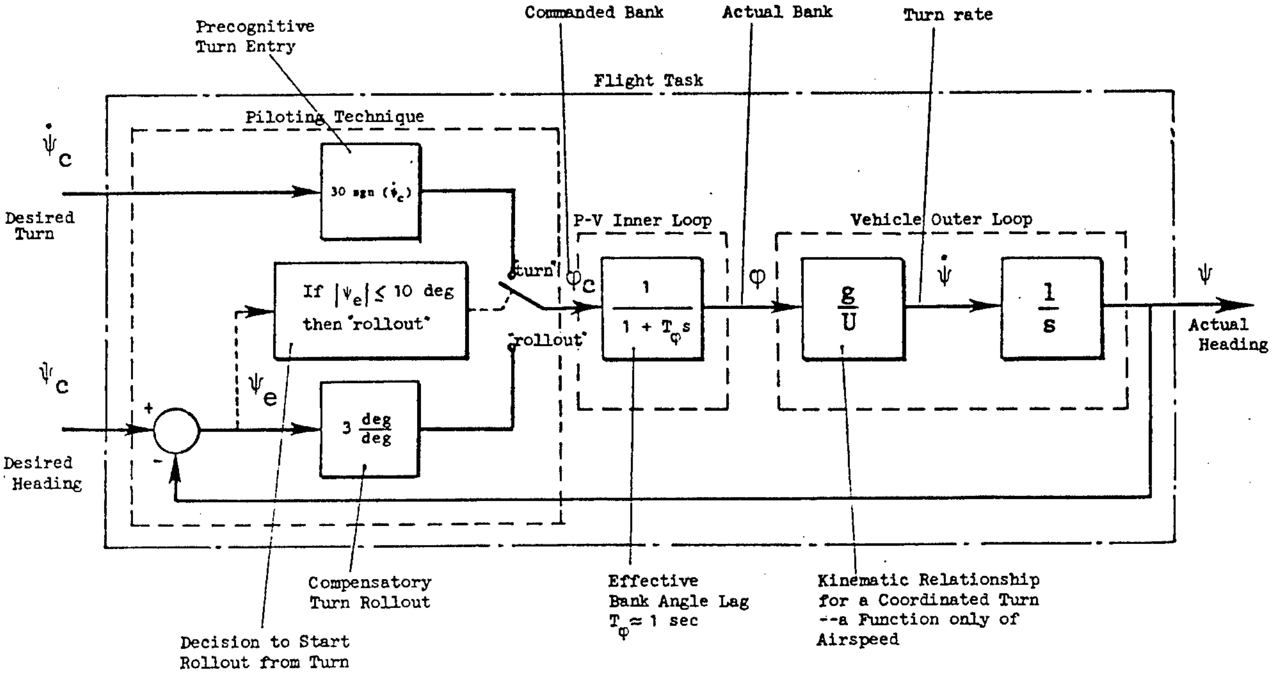


Figure 9. Heading Change Task and Piloting Technique

Verbal Description (Ref. 3)

In a typical approach, the pitch attitude is 3 to 4 deg for the DC-10. As speed is decreased to threshold speed, the pitch attitude will increase about 1 deg. Landing flare is normally initiated at approximately 30 to 40 ft above the runway surface. In the hypothetical case, if the airplane is flared to a zero rate of descent with idle thrust and a speed of just under threshold at touchdown, the pitch attitude will be 8 to 9 deg. However, with a typical low rate of descent at touchdown, the pitch attitude will normally be 7 to 8 deg. Landing with a 50 flap setting decreases the pitch attitude approximately 1 deg over that for 35 flap. There is ample tail ground clearance for a normal 35 to 50 flap approach and landing, even with the main landing gear struts fully compressed and flat tires. Fuselage contact with the runway will not occur until approximately 14 deg pitch attitude.

Feedback Control Description (Ref. 4)

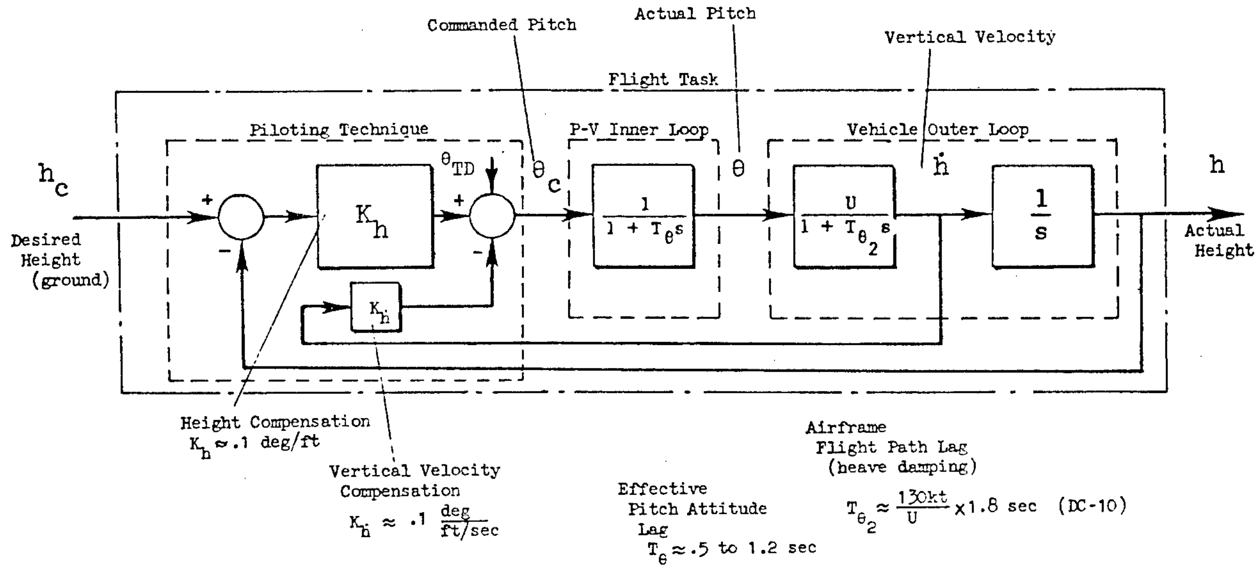


Figure 10. Landing Maneuver Task and Piloting Technique

Verbal Description

Effect a speed change Maneuver by simultaneously changing pitch attitude and offsetting flight path upset by suitable use of collective control. Stabilize on the desired new speed through appropriate use of pitch attitude.

Feedback Control Description

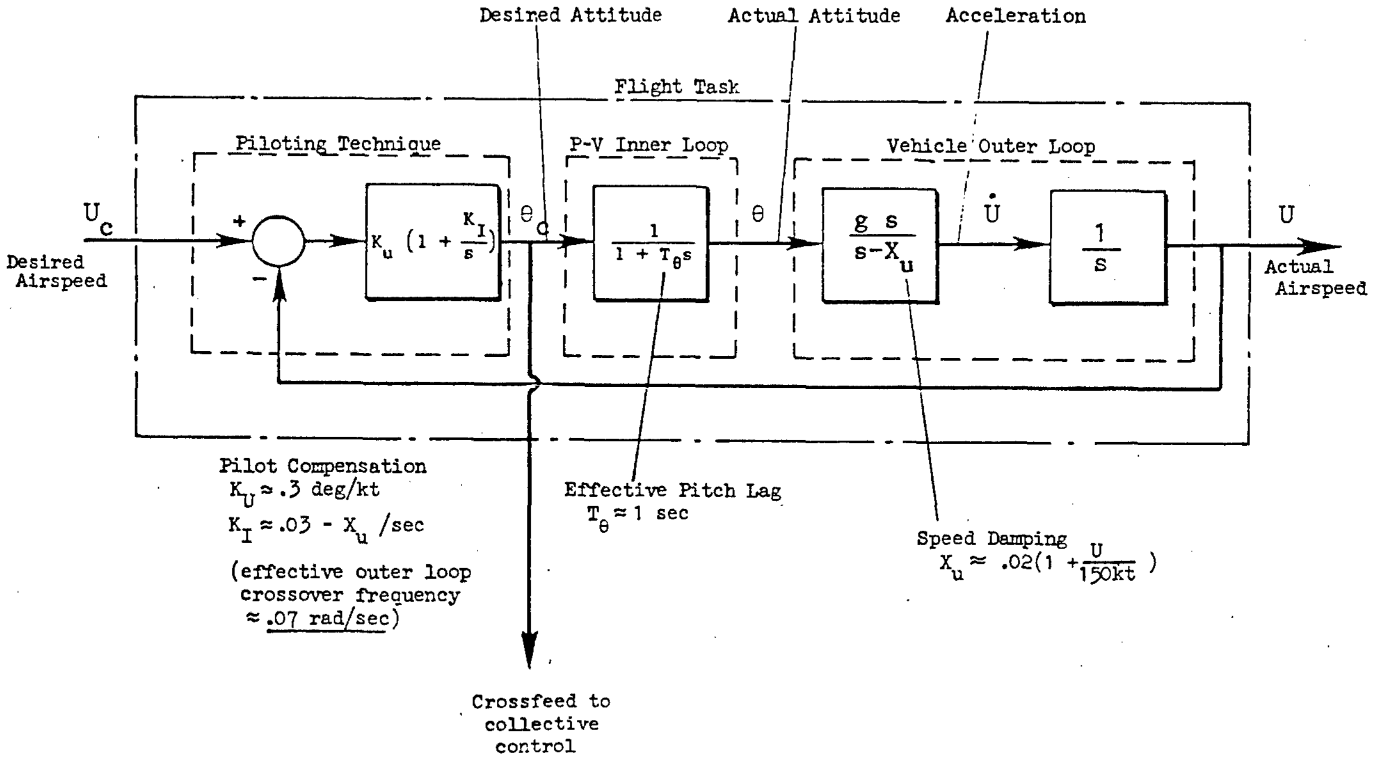


Figure 11. Up-And-Away Acceleration/Deceleration Task and Piloting Technique

Verbal Description

Fly a descending, decelerating visual approach terminating in a 40 ft hover over a landing pad. Avoid abrupt maneuvers. No approach guidance includes only standard aircraft instruments normally used for visual approaches.

Feedback Control Description (Ref. 5)

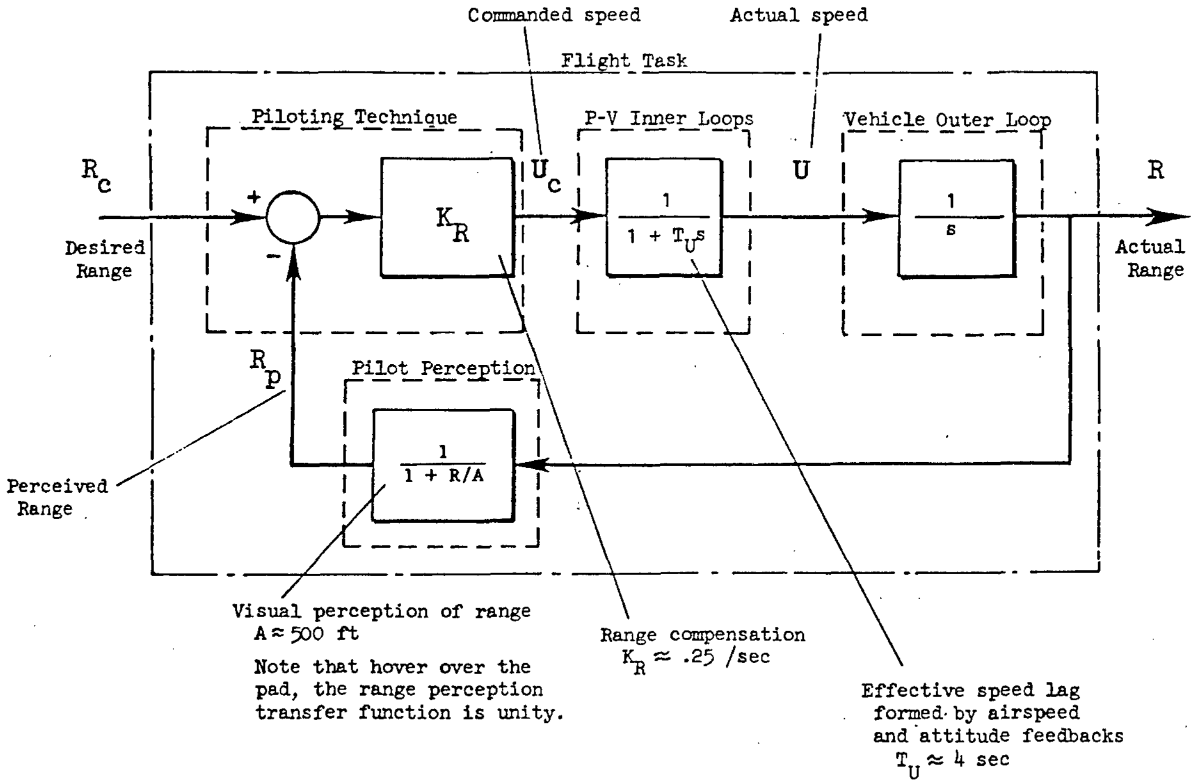
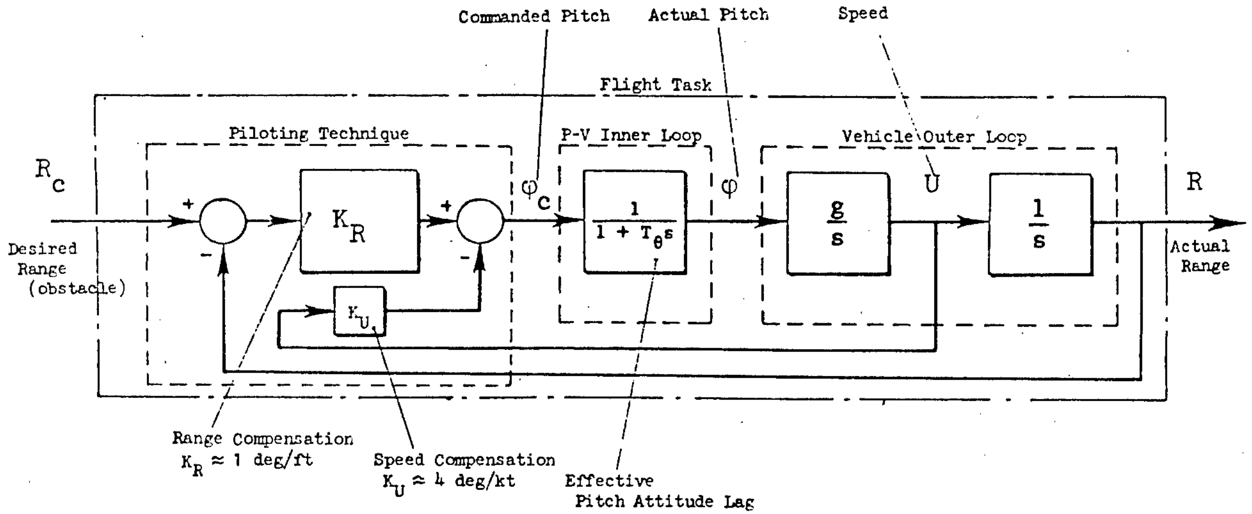


Figure 12. Decelerating Approach to Hover Task and Piloting Technique

Verbal Description

Halt all forward motion with respect to the terrain as rapidly as possible without ground contact (of the tail rotor) or excessive increase in height.

Feedback Control Description



(Involves precise control of attitude and is critical to success of maneuver.)

Though not shown, the piloting technique for controlling height is as crucial as the deceleration per se.

Figure 13. NOE Quickstop Deceleration Task and Piloting Technique

relationships are given in Table 1. Additionally, we are trying to apply such relationships to several diverse areas including evaluation of simulator fidelity. We are also attempting to apply non-intrusive pilot measurements to improve understanding and specification of handling qualities; to improve measurement and quantification of level of skill development in aircrew training; and, finally, to improve thereby our understanding of the sources and causes of aircrew error, which compromise flight safety.

TOOLS AND CONCEPTS BEING APPLIED

To provide a better idea of the non-intrusive techniques that we are trying to develop and use, we would like to present a list of tools and concepts which are used in non-intrusive piloting technique assessment.

Manual Control Theory

First, and perhaps most important, are the basic ideas of manual control theory which have been developed over a number of years to represent observed pilot behavioral changes due to training (learning effects) and variations in task variables (adaptation). It is important to note that we are not treating the pilot behavior as somehow mysterious, random, or irrational; rather, we are recognizing that the pilot's behavior (in terms of control strategy) is defined in terms of what the pilot can perceive about the state of the aircraft and the state of the task execution and the fact that the pilot is comparing what he desires with what he perceives to be the current state of affairs. Most of these underlying ideas are conveniently summarized, for example, in Refs. 6 and 7 together with their extensive bibliographies. They offer a practical basis for describing the important characteristics of the human operator's perceptually-centered control strategy, a concept which we shall define next.

Perceptually-Centered Control Strategy

In performing a multioperator/multiloop control task (such as controlling traffic or flying an airplane) each human operator's control actions are based on his perception of the situation to be controlled. The perception involves a selection among available external or internal cues in more than one sensory modality, and the verbal or manual control actions can be thought of as deriving from cognitive and psychomotor transformations of these perceptions. It is this selection plus transformation process that we will define as the operator's control strategy. In essence, the selection/transformation process generates the control output.

Operational manual control systems are typically designed to require far less than the operator's ordinary limiting capabilities in adaptation* or learning* of control strategy. Performance decrements due to environmental stress are therefore seldom observed except in emergencies when the operator is near his limiting performance. Intrinsic skill development limits in control strategy can be measured only under high task-induced stress conditions. Particular control tasks can be designed to emphasize particular skill factors in accordance with the successive organization of perception (SOP) theory, another important concept first set forth in Ref. 8 and deserving of brief elaboration here.

The Successive Organization of Perception Theory for Skill Development

The SOP theory describes the human operator's synthesis, by means of internal organizational modifications derived from training and experience, of progressive arrangements of control strategy from compensatory through pursuit to precognitive stages —— or, in other words, a

---

*Adaptation changes strategy and performance in a new environment, whereas learning or skill development changes strategy and performance in successive encounters with the same environment.

progression from behavior patterns which exhibit closed-loop, to combined open- and closed-loop, to purely open-loop properties.

The SOP theory leads to an understanding of both progressive and regressive control and monitoring behavior during training, transfer, rehearsal, and stressful operations. It can also be associated with at least one concept of perceptual motor loading.

In our concept of the perceptual-motor loading components of pilot workload, perceptual-motor activity is carefully defined to involve only conscious perceptions and actions. It is handling the unpredictable (emergency) or unfamiliar (lack of practice) which taxes the operator's workload capacity. In this context the three stages of SOP can be compared on a perceptual-motor load (PML) basis.

1. Initial stage (compensatory control). The early phases of learning predominantly involve continuous, conscious activity. We would, therefore, expect a high PML during compensatory operations.

2. Intermediate stage (pursuit control). The operator makes use of any coherence, pattern, or discrete cues in the presented stimuli. Compensatory control activity, although present, experimentally shows a regression. This implies a lower sensory-motor activity level. Therefore we would expect the pursuit level of operation to have a lower PML than the compensatory stage..

3. Final stage (precognitive control). At this level of skill, most of the operator's output consists of execution of stored commands, and his conscious perceptual activity is mainly concerned with decision-making activity. This should result in a lower PML for a given control task.

Pilots indicate that one effect of noncurrency is a general roughness of control application and lack of precision. This causes them to spend more time controlling the aircraft (higher workload), which leaves less time for other procedural matters involved in complex tasks. This degradation of control skill corresponds to regression on the SOP control skill scale given above. Thus lack of practice on a skill increases the perceptual motor loading of that skill, resulting in less workload reserve capacity for other elements of a complex task. It is apparent that lack of practice could reduce this capacity to less than that required for carrying out the remaining elements of a complex task; or a simple emergency could arise that would consume additional capacity, thus overloading the pilot and resulting in degraded system performance, if not failure.

The compensatory-pursuit-precognitive pathways structure is suitable to represent not only a pilot or controller's progression to, or regression from, higher levels of internal cognitive system organization in a given situation, but also to represent grossly the possible loop structures when different levels of display information are provided. In addition, the process can even describe the procedural organization and operating discipline among individuals on the flight deck or within the air traffic control system. Assessment of pilot dynamics and workload thus requires special attention to the SOP and the pilot control strategy forms which distinguish each stage of the SOP. For a discussion of this, the reader is directed to Ref. 9.

Control Strategy Models

Next we recognize the concept and the distinction between continuous and discrete maneuver control strategy models. The basic distinction between a continuous model and a discrete maneuver model is that the continuous model is a basic regulatory or tracking task. The discrete maneuver is something that the pilot does once. This could be a landing flare; it could be responding to a traffic controller with a change of heading, a change of altitude; or it could be a high-g air combat maneuver.

Basic and applied research going back more than two decades has demonstrated that the human operator's control strategy can be fairly accurately described by linear differential and/or difference equations. When the control task is time-varying, such as controlling an aircraft on final approach, then the coefficients of the differential equations will, in general, also be time-varying. The differential equations describe the human's operations on his perceptions, and it is these operations that characterize his control strategy. A convenient (and widely accepted) method of expressing these differential equations is to use the Laplace transform or the Fourier transform.

It is essential that the analyst choose a likely candidate for a control strategy model in order for the subsequently estimated parameters of the pilot's or controller's describing function (a particular type of Fourier transform) to be valid. Indeed it is the analyst's exercise of a priori knowledge and experience in choosing the form of the model which permits the use of a very simple and efficient mathematical algorithm for performing the estimation.

Effective values for pilot parameters including gains, lead time constants, and delays can be approximated from the identified frequency response of the pilot. One significant feature of this kind of information is that where the pilot is anticipating visually perceived information, i.e., leading it, he is faced with increasing workload. Likewise increased values of pilot gain alone will reflect increasing mental concentration on the control task itself. In addition, the pilot may be faced with an increased workload requiring division of his attention betweeen controlling and monitoring tasks, either within the cockpit or outside the cockpit upon acquisition of contact flight cues. Decreased values of gain and increased pilot remnant, sometimes accompanied by increased values of effective neuromuscular delay may reflect division of attention, particularly if substantiated by accompanying eye-point-of-regard measurements.

When other modalities are employed to reinforce the visual modality, such as rotary motion cues from an actual aircraft, certain of the visual workload requirements can be reduced. In the case of rotary motions greater than semicircular canal threshold levels, the low-frequency visual lead generation requirements are reduced. In essence, the rotary motion cues permit the pilot to close an inner loop akin to that of a rate gyro. The net effect is to reduce the effective whole-task time delay by about 0.15 sec (which also happens to be the effective neuromuscular delay increment required to develop a first-order low-frequency lead). Thus the total visual workload will have been reduced by the sensing of motion. One could say that the total workload has not been changed because what had previously been done with the visual channel was now accomplished by the visual plus motion channels acting together. However the motion loops are essentially autonomic — nearly reflexive in nature — so the additional modality reduces the conscious workload as reflected by pilot ratings. The reinforcing use of aural and other modalities also has a qualitatively similar effect.

Alternative Domains for Identification of The Human Operator's Control Strategy

Another important tool for identifying control strategy is simply depicting the pilot-vehicle-task behavior in a variety of domains: time domain, frequency domain, and the phase-plane domain. Each provides a certain kind of insight into how the pilot is operating and the perceptual dynamics involved. Both the time and frequency domains enjoy wide use. The phase-plane involves suppression of the time domain and shows the direct relationship between one state variable or control with another state variable or control. The classical kind of phase-plane diagram is a plot of a velocity versus a displacement. We extend the idea of the phase plane to include a plot of, for example, control displacement versus a variable being controlled.

Running Least Squares Estimation Methods

One final, very important computational tool or technique is the use of running least-squares estimation methods. These methods can provide us with a basic automatic parameter identification capability but at very low cost in terms of software complexity, computer speed, or computer capacity; and they permit on-line results.

The non-intrusive techniques employed for identification of the human operator's control strategy from flight and simulation test data can be classified as sampled data correlation in both the time domain and the phase plane using a least squares criterion with existing inputs. A running multiple linear regression technique called NIPIP* has been used to estimate the human operator's control strategy on-line in real time as well as after completion of the testing. A "sliding" time window is used to refresh the data; thus time-varying control strategy can be identified. Sampled data correlation in the phase plane has also been employed successfully.

SOME LESSONS LEARNED

We would now like to mention some of the important lessons learned in applying non-intrusive measurement procedures. First, we believe that it is crucial to exploit fully simple, easily visualized mathematical models before increasing model complexity. The pilot-vehicle should be studied one loop or one axis at a time. Frequency partitioning of tasks is highly recommended. We have also found that it is prudent to rely heavily on first-principles models of the vehicle, the pilot, and the task. We take advantage of the principles of physics, the theory of manual control, the rules for automatic control design and adjustment if necessary (e.g., Ref. 10), and the geometrical relationships which describe the world which the pilot perceives and the flight paths which the aircraft flies. These models provide the analyst with the basic insight for adopting increased complexity if that is necessary. We have generally found, however, that, after studying behavior in terms of simple mathematical models, we sometimes are able to recast our results in even simpler, more cogent terms. In contrast, complex computer analysis routines and reliance on large-scale computers sometimes actually reduce the number of things which can

---

*Non-Intrusive Pilot Identification Procedure.

be analyzed and frequently tend to obscure the basic insights and understanding of events.

In order to understand better the results from the measurement procedures, it is helpful, if possible, to work in the running time domain; that is, to observe the pilot estimation results as they are happening and to permit non-stationarity in these results. It is important, for example, to allow for alterations in pilot control strategy over the period of a landing approach. It is therefore important to have a computational procedure which is fast enough to identify non-stationary piloting technique. Long-term averaging of pilot measurements is valid only when one is sure that the flight task is not changing in the long term and that the pilot has not changed the piloting technique due to environmental disturbances.

It is especially important when interpreting results to consider the various perspectives of the piloting, engineering, psychological, and training disciplines. No single one of these disciplines dominates assessments that are being made of the pilot dynamics or the pilot workload. At the same time, there are perspectives, technical approaches, and bodies of literature within each discipline which may trigger additional insight.

Next we have found that it is important to treat mission-oriented flight tasks as opposed to the more artificial and contrived tasks which may be conducted in a laboratory environment. There is no question that laboratory results have been invaluable in establishing basic ideas of manual control theory and relationships between pilot dynamics and pilot workload, but laboratory test scenarios have limited applicability. Mission-oriented flight tasks require that we deal in a realistic context, with multiple control loops and with the proper mix of cognitive and psychomotor tasks. In so doing, however, we are compelled to develop and apply non-intrusive measurement techniques.

Finally, we have found that it is important to be skeptical of extrapolating simulator results to flight. For a given task, and especially for critical, difficult tasks, one should attempt to start with flight data in order to validate a simulator. Even on costly, very sophisticated large amplitude motion simulators, we have found that in performing such flight tasks as landing flare or low-altitude helicopter maneuvering, the pilot does not necessarily fly the simulator with the same strategy as the pilot, in fact, flies the real aircraft. Measurements of control strategy and perceptual dynamics, which have been made non-intrusively, bear this out.

RECENT AND ONGOING PROJECTS

We conclude by mentioning some recent and on-going projects which involve non-intrusive piloting technique measurement procedures. These projects have spanned several areas and have included various applications for measurement of pilot control strategy.

Evaluation of Powered-Lift STOL Safety Margin Displays

One general application has been the documentation of how pilots use special displays, for example, use of a safety margin display for the operation of a powered-lift STOL aircraft (Ref. 11). This work was performed for NASA and involved use of a fixed-base simulation of the Augmentor Wing Jet STOL Research Aircraft. Non-intrusive measurements revealed tight tracking for one display configuration but relatively loose tracking for another. The latter configuration produced worse pilot ratings, a predictable result with knowledge of the system dynamics and the piloting technique of the pilot.

Identification of Pilot Control Strategy in the Use of Head-Up Displays for CTOL Aircraft

NIPIP was used to identify pilot control strategy online in real time during a ground-based simulation of two competing concepts for head-up displays (HUD) and a head-down attitude director indicator for use in conventional takeoff and landing (CTOL) aircraft (Ref. 12 and 13). Eleven professional airline captains participated. Data for three of the pilots were selected for more detailed comparison and discussion.

The "flight path" HUD was flown with a higher bandwidth (or crossover frequency) and lower phase margin than was the "flight director" HUD. The "flight director" HUD provided a fixed aircraft symbol and a moving dot to display the vertical and lateral guidance director commands together with a numerical display of speed. Other status information had to be derived by the pilot from the external visual field. The "flight path" HUD offered a conformal symbolic display of the direction of the aircraft's flight path, suitably compensated with lead equalization, the reference glide slope angle, and an angle proportional to the glide slope error. In addition, the pilot could perceive an effective director command in the symbology of the "flight path" HUD by observing the angular difference between compensated aircraft flight path and glide slope angle. The "flight path" HUD also provided an analog "fast-slow" speed signal. Both HUDs incorporated artificial horizons.

Examination of the describing function data revealed that the pilots were able to fly the head-up displays with a fairly high gain and use lead compensation in their control strategy, but the describing functions for the head-down display exhibited lag compensation and a low bandwidth.

The pilot's describing functions exhibited a trend in time-varying behavior when using the "flight director" HUD and head-down display but to a much less degree when using the "flight path" HUD. Usually the pilots increased their gain and decreased their control latency as they came closer to the minimum decision altitude.

NIPIP was also able to detect control strategy errors with the head-down display and the "flight path" HUD. These errors were due to control reversals with the head-down flight director and to accidentally using the wrong element in the display with the "flight path" HUD. After "breaking out" (i.e., beneath the cloud cover), the pilot apparently became confused as to which symbol to track in the display. This caused an instability in the flight director loop. This sudden change in control strategy caused a sharp decrease in the amplitude and phase of the pilot describing function and corresponding decreases in crossover frequency and phase margin. The unstable condition lasted for only a few seconds, but NIPIP identified the control strategy error before the pilot was aware of it and applied the proper control action.

Design of a Simulator to Study Human Error

A study was conducted for the Man-Machine Systems Division of NASA Ames Research Center which focused on human error. The study resulted in the description of a theory of human error (Ref. 9), how human error might be objectively measured (Ref. 14), and recommendations for the design of a simulator facility for making such measurements (Ref. 15). The role of non-intrusive measurements was considered essential in this application.

A Comparison of Piloting Technique Made in an Airline Training Simulator and in Actual Flight

An analysis of pilot control strategy, both from an airline training simulator and an actual DC-10, was performed for the landing maneuver (Ref. 4). An emphasis was placed on identifying useful metrics, quantifying piloting technique, and defining simulator fidelity. On the basis of DC-10 flight and simulator measurements recorded in about 200 landings by 32 pilots — 13 flight-trained and the remainder simulator-trained — a revised model of the landing flare was hypothesized which accounted for reduction of sink rate and preference for touchdown point along the runway. The flare maneuver and touchdown point adjustment could be described by a pitch attitude command pilot guidance law consisting of altitude and vertical velocity feedbacks. The pilot gains which were identified directly from the flight and simulator data showed that the flare was being executed differently in each medium. In flight most of the subject pilots exhibited a significant vertical velocity feedback which was essential for well-controlled sink rate reduction at the desired level of response (bandwidth). In the simulator, however, the vertical velocity feedback appeared ineffectual and led to substantially inferior landing performance. The absence of the vertical velocity feedback implied a simulator fidelity problem with a potential impact on pilot workload, and several specific possibilities were discussed. The pilot model of the maneuver provided insight into which aircraft types could be simulated without incurring the apparent fidelity limitation encountered in this case.

Identification of Pilot Control Strategy in Simulated Approaches to an Aircraft Carrier

An issue of concern is the influence of pilot-generated noise on the validity of pilot control strategy identified from test data. One recent study resulted in predicting results of a simulation of the carrier aircraft pilot's approach control strategy in the presence of pilot remnant (Ref. 16). The aircraft dynamics and the turbulence environment were representative of a (T-2C) trainer-type aircraft. NIPIP was used to identify the pilot's control strategy required by this highly-coupled, multiloop control task. Of particular interest was whether the pilot learns to adopt a workload-reducing pursuit crossfeed of throttle-to-pitch attitude command which enables him to achieve adequate approach precision. The results were presented in terms of frequency responses of the individual elements of the pilot's control strategy. It was found that NIPIP can identify the pilot's describing functions even in the presence of rms pilot remnant amounting to 25 percent of the observed rms states when no remnant is present. The next step is to apply NIPIP to a real time, piloted simulation of the same control task. This is planned for the Visual Technology Research Simulator at the Naval Training Equipment Center in Orlando, Florida.

Determination of Motion and Visual Systems Requirements for Army Training Simulators

A study was conducted for the Army Research Institute directed at identifying simulator fidelity requirements for training (Ref. 1). Both rotary- and fixed-wing Army missions were considered. The result of the study was the description of a systematic method for approaching simulator fidelity requirements which included direct measurement and cataloging of the pilot control strategy and perceptual dynamics involved in each training objective (piloting task). Such measurements would be required both in flight and in existing simulator facilities. An analytic approach was then described which, based on data obtained, could be used to predict the utility of future systems or to set system requirements. Non-intrusive pilot measurements were thus considered necessary for exploring how pilots operate the actual aircraft as well as how they operate the training simulator equipment.

Design and Operation of a New Navy Advanced Trainer

A study was conducted for the Northrop/Vought VTXTS design team which involved non-intrusive pilot measurements for evaluating simulator hardware (Ref. 17). A carrier approach task was flown; and the pilot control strategy was measured relative to vertical flight path control, pitch attitude control, and throttle-to-attitude crossfeed. Several pilots with varying backgrounds and levels of skill were used, and it was possible to distinguish these in the control strategy measurements.

Research Simulator Fidelity Using Flight and Simulator Measurements of Control Strategy

Currently the fidelity of a UH-60 Black Hawk helicopter simulation at NASA Ames Research Center is being evaluated (Ref. 18). This is being accomplished using pilot control strategy measurements from actual UH-60 flights for a variety of Army flight tasks including the NOE regime. Corresponding control strategy measurements will be performed on the Ames Research Center Vertical Motion Simulator (VMS) in order to find differences in the use of essential motion and visual cues and, thereby, to infer simulator fidelity. This work will also result in an initial cataloging of several basic flight maneuvers in terms of pilot control strategy and the overall closed-loop dynamics of each task.

CONCLUSION

Non-intrusive assessment of pilot dynamics and pilot workload has become increasingly more important as we begin to focus on real-world piloting tasks in actual flight rather than in simulated flight. Maneuvers such as landing flare, tracking tasks such as the carrier approach, and operating environments such as Army NOE often preclude intrusive pilot measurements or contrived laboratory experiments. Knowledge of the piloting technique involved in such tasks or environments reveals much in terms of essential cues for simulation, handling qualities requirements, and training needs.

The recent and on-going studies mentioned herein indicate the wide spectrum of applications already considered for non-intrusive measurements. During these same projects, much has been accomplished in developing and using measurement techniques. It has been important to exploit many different tools and concepts within the areas of manual control theory, vehicle dynamics, control system analysis, and numerical analysis. In so doing, however, effective management of analysis techniques is crucial. This has tended to force an emphasis on relatively simple modeling and measurement procedures. At the same time, mathematical simplicity has proved highly effective in gaining understanding and in finding where increased complexity is needed.

REFERENCES

1. Heffley, Robert K., Warren F. Clement, Robert F. Ringland, et al., Determination of Motion and Visual System Requirements for Flight Training Simulators, Systems Technology, Inc., Technical Report No. 1162-1, August 1981 (forthcoming ARI report).

2. Anon., Instrument Flying, Air Force Manual, AFM 51-37, 15 August 1979, pp. 2-13.

3. Anon., United Airlines, DC-10 Flight Manual, Training and Reference, 1973-1974, United Airlines Flight Training Center, Stapleton International Airport, Denver, Colorado, October 17, 1977, pp. 4-16.

4. Heffley, Robert K., Ted M. Schulman, Robert J. Randle, Jr., and Warren F. Clement, An Analysis of Airline Landing Data Based on Flight and Training Simulator Measurements, Systems Technology, Inc., Technical Report No. 1172-1, 1981 (forthcoming NASA Technical Memorandum).

5. Heffley, Robert K., "A Model for Manual Decelerating Approaches to Hover," Proceedings of the Fifteenth Annual Conference on Manual Control, AFFDL-TR-79-3134, 1979, pp. 545-554.

6. McRuer, D. T., and D. H. Weir, "Theory of Manual Vehicular Control," Ergon., Vol. 12, No. 4, July 1969, pp. 599-633; and Trans. IEEE, Vol. MMS-10, No. 4, Pt. II, December 1969, pp. 257-291.

7. McRuer, D. T., and E. S. Krendel, Mathematical Models of Human Pilot Behavior, AGARD-AG-188, January 1974.

8. Krendel, Ezra S., and Duane T. McRuer, "Servomechanisms Approach to Skill Development," J. Franklin Inst., Vol. 269, No. 1, January 1960, pp. 24-42.

9. 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 (forthcoming NASA CR).

10. McRuer, Duane, Irving Ashkenas, and Dunstan Graham, Aircraft Dynamics and Automatic Control, Princeton University Press, Princeton, New Jersey, 1973.

11. Heffley, Robert K., and Wayne F. Jewell, *Study of a Safety Margin System for Powered-Lift STOL Aircraft*, NASA CR-152139, May 1978.

12. 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.

13. Jewell, Wayne F., and Ted M. Schulman, *A Pilot Control Strategy Identification Technique for Use in Multiloop Control Tasks*, Systems Technology, Inc., Technical Report No. 1153-2, August 1980.

14. Clement, Warren F., Robert K. Heffley, Wayne F. Jewell, and Duane T. McRuer, *Technical Approaches for Measurement of Human Errors*, Systems Technology, Inc., Technical Report No. 1156-2, July 1980 (forthcoming NASA CR).

15. Clement, Warren F., R. Wade Allen, Robert K. Heffley, et al., *Functional Requirements for the Man-Vehicle Systems Research Facility*, Systems Technology, Inc., Technical Report No. 1156-3, October 1980 (forthcoming NASA CR).

16. Jewell, W. F., H. R. Jex, R. E. Magdaleno, and R. F. Ringland, *Reports by Systems Technology, Inc., in Support of Carrier-Landing Research in the Visual Technology Research Simulator*, NAVTRAEQUIPCEN 78-C-0060-10, December 1981.

17. Heffley, Robert K., and Samuel J. Craig, *The Use of Pilot Control Strategy Identification Techniques to Measure Simulator Fidelity for the VTXTS*, to be presented at the SSES/SFTF Symposium, "Simulation-Aircraft Test and Evaluation," March 16 and 17, 1982, Naval Air Test Center, Patuxent River, MD.

18. Cleveland, William B., Robert K. Heffley, and David L. Key, *Flight Measurements of Pilot Performance and Workload for UH-60 Simulation Validation Assessment*, to be presented at the AIAA Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, January 19-21, 1982, Air Force Flight Test Center, Edwards Air Force Base, CA.

The Optimal Control Model for the Human Operator: Theory, Validation, and Application

by

William H. Levison
Bolt Beranek and Newman, Inc.
Cambridge, MA 02238

Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics

January 19-21, 1982
Edwards Air Force Base, CA

ABSTRACT

The Optimal Control Model for the Human Operator is reviewed. Underlying concepts are presented and the model structure is described. Special emphasis is given to the treatment of attentional workload, including the relationship between workload and closed-loop performance. Validating experimental results are presented, and both predictive and diagnostic applications of the model are reviewed. Areas of further model development and application are summarized.

INTRODUCTION

The Optimal Control Model (OCM) for the human operator is based on the assumption that the well-motivated, well-trained human operator will act in a near optimal manner subject to the operator's internal limitations and understanding of the task. This assumption is consistent with notions of human response behavior discussed in the psychological literature. What differentiates the OCM from other models of the human operator are the methods used to represent human limitations, the inclusion in the model of elements that compensate optimally for these limitations, and the extensive use of state-space concepts and the techniques of modern control theory.

Clearly, if the basic optimality assumption is to yield good results, it is necessary to have reliable, accurate and meaningful models for human limitations. Insofar as possible, these models (or their parameters) should reflect intrinsic human limitations or should depend primarily on the interaction of the operator with the environment and not on the specifics of the control task. It is also desirable that the description of human limitations involve as few parameters as possible and that it be commensurate with the modern control system framework that is being employed. These principles have guided the development of the models for human limitations that will be described below.

MODEL DESCRIPTION

The objective of this section of the paper is to familiarize the reader with the basic structure of the OCM and to provide an appreciation of the types of assumptions and considerations required to apply the model. Discussion is largely verbal and conceptual; the reader is referred to the literature for mathematical details [1,2].

In order to apply the OCM, the following features of the environment must be specified: 1) a linearized state variable representation or model of the system being controlled; 2) a stochastic or deterministic representation of the driving function or environmental disturbances over which the operator must exert control; 3) a linearized "display vector" summarizing the sensory information utilized by the operator (including visual, vestibular and other sources as appropriate); and 4) a quantitative statement of the criterion or performance index for assessing operator/machine performance. The specific assumptions concerning this description that are necessary to apply the theory are given in reference 1.

The OCM, diagrammed in Figure 1, is a model for the dynamical response behavior of the closed-loop control system. Because the model is capable of treating multi-variable systems, all system variables shown in the figure are represented as vector quantities. The portion of the model structure designated as "Human Operator Model" contains elements related to the operator's adaptive response behavior and to limitations that constrain this behavior. We briefly review these model elements in the order corresponding to the flow of information indicated in the figure.

First, the displayed variables are assumed to be corrupted by "observation noise" introduced by the human operator. This noise is analogous to the internal noise level postulated in signal detection theory and, as discussed later, provides one means by which the model accounts for human limitations in perceptual resolution, central-processing, and attention-sharing capacity.

At this point the model is dealing with a noisy representation of the displayed quantities. That representation is then delayed by an amount, τ, representing internal human processing delays.

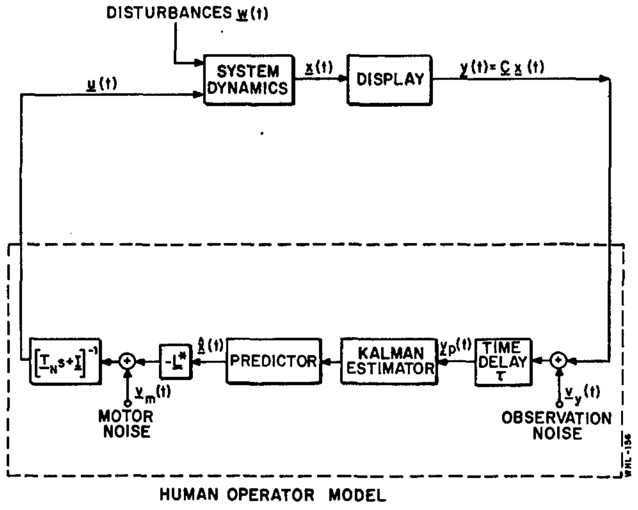


Figure 1. Structure of the Optimal Control Model

In theory, it should be possible to consider differential delays for different sensory channels, but this additional complication has generally been unnecessary in past model applications to manual control data.

The central elements of the model are represented in the blocks described as the Kalman estimator and predictor. Their purpose is to generate the best estimate of the current state of system variables, based on the noisy, delayed perceptual information available. These elements compute the estimate of this state so as to minimize the residual estimation uncertainty; they represent the operator's ability to construct from his understanding of the system and his incomplete knowledge of the moment-by-moment state of the system, a set of expectancies concerning the system behavior at the next moment in time. These elements reflect the assumption that the human operator has both an internal model of the dynamics of the system being controlled, and a representation of the statistics of the disturbance driving the system.

Given the best estimate of the current system state, the next model element (labeled L*) assigns a set of control gains or weighting factors to the elements of the estimated state, in order to produce control actions that will minimize the defined performance criterion. As might be expected, the particular choice of the performance criterion determines the weighting factors, and thus the effective control law gains.

Just as an observation noise is postulated to account for perceptual and central processing inadequacies, a motor noise is introduced to account for an inability to generate noise-free control actions. In many applications this noise level is insignificant in comparison to the observation noise, but where very precise control is important to the conditions being analyzed, motor noise can assume greater significance in the model. Finally, the noisy control response is assumed to be smoothed by a filter that accounts for an operator bandwidth constraint. In the model, this constraint arises directly as a result of a penalty on excessive control rates included in the performance criterion. The constraint may mimic actual physiological constraints of the neuromotor system or it may reflect subjective limitations imposed by the pilot.

This, then, provides a conceptual description of the elements of the optimal control model. It should be emphasized that the parameter values that must be provided by the investigator correspond to the human limitations that constrain behavior. With these limitations as the constraints within which performance is produced, the model predicts the best that the operator can do. A large backlog of empirical research provides the data necessary to make realistic estimates of the appropriate parameter settings in the manual control context. This research has shown that, for an important class of control tasks, these parameters are relatively invariant with respect to changes in task environment, thus enhancing the model's predictive capability [1-4].

Let us now review in more detail certain critical features of the model.

Task Requirements

The pilot is assumed to adopt a control strategy that minimizes a quadratic performance index of the following form:

$$J=E\{\lim_{T\to\infty} \frac{1}{T} \int_0^T [\sum_i q_i y_i^2(t) + \sum_i (r_i u^2(t)+g_i \mathring{u}^2(t))]dt\} \qquad (1)$$

where "q", "r", and "g" are the cost coefficients associated with display, control, and control-rate variables, and E signifies statistical expectation. Single-variable laboratory tracking tasks can often be modeled under the assumption that the operator minimizes a weighted sum of mean-squared error and mean-squared control rate [1-5]. For more complex tasks, cost coefficients will be associated with more than one element of the display vector. When analyzing aircraft approach-to-landing, for example, penalties are generally associated with path, attitude, control displacement and control rate. Reasonable model predictions have been obtained on the basis of assumed maximum allowable values for various control and response variables [6,7].

The performance index of Eq(1) leads to a constant set of control gains because of the assumption that "cost" is computed over an arbitrarily long time. Although most model applications have been based on this assumption, this is not a theoretical limitation of the OCM, and tasks (such as airplane flare and touchdown) involving "terminal cost" and time-varying control gains can be treated.

Perceptual Environment

The perceptual submodel contained in the OCM is represented by the following relationships:

$$\underline{y}(t) = \underline{Cx}(t) \qquad (2)$$

$$\underline{y}_p(t) = \underline{y}(t) + \underline{v}(t) \qquad (3)$$

where $\underline{x}(t)$ is the vector of system "states", $\underline{y}(t)$ is the display vector, $\underline{y}_p(t)$ an observation noise vector, and $\underline{C}$ a linear transformation matrix.

As indicated by these equations, each display variable to the operator is assumed to be a linear combination of system states and is further assumed to be corrupted by a white noise process. Each noise process is assumed to be linearly independent of other such noise processes and of external inputs acting on the system.

Because the OCM is an informational model of the task environment (including the human operator), the perceptual environment should be analyzed to determine both its informational content relevant to the specific task, and the quality of such information in terms of allowing the operator to improve his estimates of system state variables. Informational content is reflected in the composition of the display vector $\underline{y}$. This vector should include each potentially relevant cue determined from analysis of the display environment and described as a linear combination of system state variables. The question of whether such a cue is of sufficient quality to be useful in a particular control situation is a separate consideration as discussed below.

Identification of relevant perceptual cues is straightforward in some cases, subtle in others. Where cues are obtained only from devices that present symbolic display information (e.g., aircraft flight-control instruments), one generally associates two components of the $\underline{y}$ vector with each moving display element: one for displacement, and one for velocity. (Independent perceptions of position and velocity are assumed). On the other hand, the set of relevant perceptual cues contained in an actual or simulated "real-world" visual scene may not be apparent; some form of visual scene analysis, such as geometric analysis, may be required to construct a suitable display vector for the OCM [8-10].

The model has two different mechanisms for treating the quality of perceptual information: (a) adjustment of the statistical properties of the observation noise, and (b) modification of the description of system dynamical response characteristics to include bandwidth limitations associated with generation and perception of displayed variables.

For an idealized perceptual environment, in which physical limitations such as threshold- and saturation-like phenomena are negligible, the covariance of each observation noise process may be considered to vary in proportion to the variance of the associated perceptual variable. A "noise/signal ratio" of around -20 dB is typical of a wide class of laboratory tracking tasks [1]. If visual or indifference thresholds are important, such as in many real or simulated flight tasks (especially those involving external visual cues), a more complex formula is used for obtaining observation noise levels [11,12]. The method employed involves a statistical threshold that results in large observation noise for signal magnitudes below the assumed threshold value. This is directly analogous to the threshold notions of signal detection theory.

Effective perceptual thresholds of 0.05 degrees visual arc and 0.2 arc degrees/second may be assumed for perception of indicator displacement and velocity, respectively, when the display consists of an indicator bar that translates with respect to a stationary reference marking [3]. The literature provides some guidance for thresholds appropriate to visual scene perception as well as whole-body motion cues [8,9,13,14]; but, for perceptual cues not heretofore modeled, some experimentation will generally be required to determine appropriate values for thresholds related to perceptual resolution. Indifference thresholds, which reflect performance requirements rather than perceptual resolution, will depend on the context of the specific tasks.

Bandwidth-related perceptual limitations may result from the properties of the physical display (e.g., inertia and damping associated with an aircraft attitude indicator) or from limitations associated with the human operator's processing of certain kinds of physical stimuli (e.g., dynamical response properties associated with perception of whole-body motion cues by the vestibular system [13,14]). Whatever their source, perceptual bandwidth limitations are expressed in linearized state-variable format and are included in the description of system response dynamics.

In summary, the following steps are required to define the OCM perceptual submodel in a specific application:

1. Analyze the perceptual environment to determine the relevant perceptual cues that may be available to the controller. These cues must be expressed as linear combinations of system state variables.

2. Modify the dynamical equations of motion of the system to include dynamical response limitations associated with perception of these cues.

3. Determine effective perceptual or indifference thresholds associated with each perceptual element, and include those threshold values in the expressions for observation noise variance.

Control Environment

The OCM has a simplified "built-in" representation of control-related limitations. First, the human controller is modeled as though there is a cost penalty associated with the rate-of-change of control effort. One mathematical consequence of this assumption is the imposition of a first-order lag network --

characterized by the "motor time constant" T_n -- associated with the operator's response strategy. Second, a Gaussian white noise process is associated with rate-of-change of each control variable.

For single-variable laboratory tracking tasks involving fast-responding systems, motor time constants on the order of 0.1 seconds have provided good model correspondence with data [1-5]. The relative invariance of T_n suggests that this parameter reflects an inherent bandwidth limitation of human controller response. For more complex and realistic tasks, however, response bandwidth is more likely to be limited by physical constraints associated with the control system [6,7].

As currently formulated, the OCM allows independent adjustment of two motor noise processes: a "driving" noise to reflect direct inputs to the control system, and an "internal" noise to degrade the model's ability to estimate certain state variables. The driving noise potentially allows the OCM to account for tremor and other inadvertent disturbances associated with execution of control, whereas the internal noise may be interpreted as a mathematical device to account partially for the operator's imperfect knowledge of system response behavior. A more thorough discussion of the treatment of motor noise may be found in Levison, Baron, and Junker [15].

Typical application of the OCM by the authors involves omitting the driving motor noise and setting the covariance of the internal motor noise to be on the order of -50 dB with respect to control-rate variance. For situations in which control resolution and operator tremor become important, one should consider a motor noise submodel parallel to that adopted for observation noise; e.g., where the noise increases (or at least levels off) as the intended control action becomes small compared to some "control threshold". Elaboration of this aspect of the OCM is an area for future research.

The OCM has been validated with data obtained largely from tasks involving nearly-isometric (i.e., force) control sticks. In such cases, the first-order filter represented by the motor time constant T_n has been sufficient to reflect bandwidth limitations that may have been imposed by the operator's neuromuscular response mechanisms. Where substantial control movement is required, however, higher-order representations are needed to account for the combined dynamical characteristics of the physical control device and the operator's neuromuscular system [16].

To treat second or higher-order representations of operator response limitations, the corresponding dynamical equations of motion are included in an expanded formulation of system response characteristics (the same treatment as accorded display-related dynamics). Some studies have been undertaken to develop detailed dynamical models for man-control interaction [17] and this remains an area for further research.

Attention-Sharing and Workload

The OCM is able to reproduce pilot response behavior in a variety of idealized laboratory tracking situations with an observation noise/signal ratio that is nearly invariant across tasks and across subject populations. The near constancy of this parameter suggests that the noise/signal ratio reflects a central processing (rather than perceptual or motor) limitation, and these results have led to the following model for central attention-sharing:

$$P_i = \frac{P_o}{f_i} \frac{1}{f_t} \qquad (4)$$

where f_t is the fraction of attention devoted to the tracking task as a whole, f_i is the subfraction of such attention devoted to display variable y_i, P_i is the noise/signal ratio associated with perception of y_i, and P_o is the baseline noise/signal ratio associated with a high-workload, single-variable tracking task (typically, -20 dB). According to this submodel, low observation noise is associated with high attentional workload.

The general nature of this article precludes a detailed review of the theoretical basis for relating attentional workload to the observation noise/signal parameter of the pilot model. Mathematical development is provided in Ref [18]. For now, let us present a number of plausibility arguments to support this notion:

1. The more the pilot suppresses his remnant (i.e., noise component of his control response), the higher his workload.

2. Suppression of random fluctuations in the pilot's response strategy requires mental effort. As shown by Levison and Kleinman [19], rapid fluctuations in the response strategy may be an important source of pilot remnant.

3. Task workload is directly related to the fraction of time spent performing the task. As shown by Levison [18], an inverse relationship exists between the fraction of time devoted to obtaining information from a display variable and the

observation noise associated with that display when scanning is rapid with respect to the signal bandwidth.

4. For information sources that are processed in parallel, the workload associated with a particular source is proportional to the information-processing "capacity" allocated to that source.

In summary, a number of information-processing attributes that we might reasonably associate with task workload can be related to pilot remnant and, therefore, through the structure of the OCM, to the noise/signal parameter given in Equation (4).

The workload submodel suggested here has been partially validated in studies of multi-variable tracking by Levison, Elkind, and Ward [4]. Wewerinke [20] has also obtained generally good agreement between subjective workload assessments and a "workload index" based partly on this submodel. Some of this validating data is presented in the following section.

The parameter f_t, representing attention to the task as a whole, may serve as a metric for task workload [6,7]; f_i, on the other hand, reflects penalties associated with sharing attention among multiple displays [4]. Attention-sharing may occur with a single modality or across modalities. The f_i may be based on available operator scanning data, or the OCM may be used to predict optimal allocation of attention (i.e., the attentional allocation that minimizes the overall performance index.)

This attention-sharing model is crucial for predicting performance in complex, multivariable tasks, and it can also serve as a basis for developing a variety of models for monitoring tasks not requiring continuous control [21]. In the applications section of this paper we demonstrate use of the OCM to predict the effects of various system parameters on the relationship between closed-loop performance and attentional workload.

VALIDATION

Basic Results

Two levels of model validity are explored: structural adequacy* and predictive validity. The former requirement is satisfied if the model structure and parameterization are sufficient to mimic experimental measurements obtained over a useful range of control tasks. A model that is structurally adequate can be used to characterize operator response behavior in terms of a few parameters (i.e., the independent model parameters) and therefore has the potential for considerable data compression. Such a model might be used, say, to quantify the effects of stress and other environmental factors on pilot response capabilities.

The more stringent property of predictive validity requires not only that the model be structurally adequate, but that operator and closed-loop system response behavior be predictable with a fixed set of independent model parameters (or a fixed set of rules for adjusting such parameters). A model of this sort can be used for system design and evaluation as well as for characterizing operator performance, as described later.

To demonstrate structural adequacy, independent model parameters were adjusted to provide the best match to data obtained from three different laboratory tracking studies which utilized simulated vehicle dynamics of approximate proportional control (K), rate control (K/s), and approximate acceleration control (K/s^2). These studies are detailed in References [23,3, and 24], respectively. The numerical search scheme described by Levison [22,25] was employed to adjust the "motor time constant" T_n, the time delay τ, the observation noise/signal ratios associated with perception of tracking error and error rate (P_e and $P_{\dot{e}}$, respectively), and the motor noise/signal ratio (P_u). Parameters providing the best joint match to variance scores and frequency response measures are given in the first three rows of Table 1.

Figure 2a shows that the linear portion of the pilot's response strategy ("gain" and "phase") as well as the stochastic

---

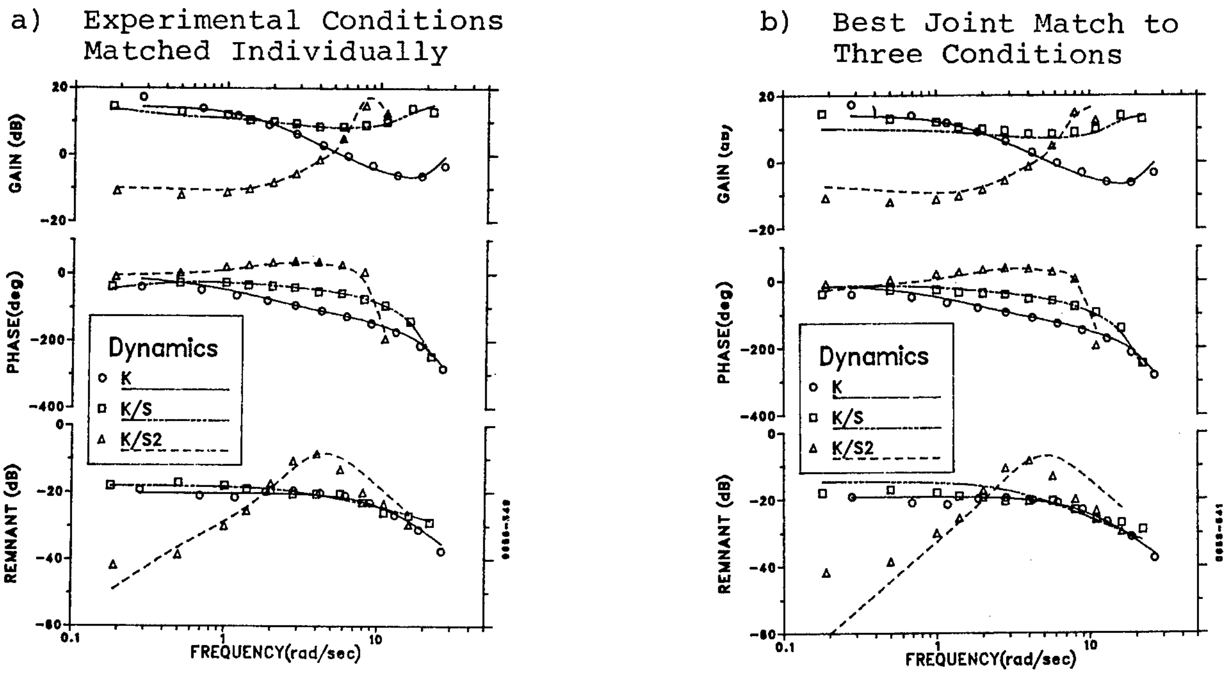
* We shy away from the term "structural validity", which would imply a close correspondence between the detailed mathematical structure of the model and the human's information processing mechanisms. Rather, we attempt to show here that the model structure is adequate to mimic input-output relationships exhibited by the data.

Table 1: Independent Model Parameters

Dynamics	P_e (dB)	$P_{\dot{e}}$ (dB)	P_u (dB)	τ (sec)	τ_n (sec)
K	-21.1	-19.5	-40.9	.167	.0812
K/s	-23.6	-18.1	-41.6	.154	.0734
K/s^2	- 4.6	-21.0	-63.9	.206	.109
(AVG)	-21.0		-50.0	0.17	.09

portion of the control response ("remnant") were matched quite closely. In addition, RMS error and control scores were all matched to within 10% of experimental values. The OCM, then, appears to have structural adequacy for the range of tasks explored.

To explore the predictive validity of the model, a single set of independent model parameters (shown in the last row of Table 1) were found to provide the best joint match to the entire body of data shown in Figure 2a. Parameterization was reduced by one degree of freedom by constraining P_e and $P_{\dot{e}}$ to have identical numerical values.



a) Experimental Conditions Matched Individually

b) Best Joint Match to Three Conditions

Figure 2. Comparison of Model and Experimental Frequency Response

Experimental results indicated by symbols, model results by smooth curves.

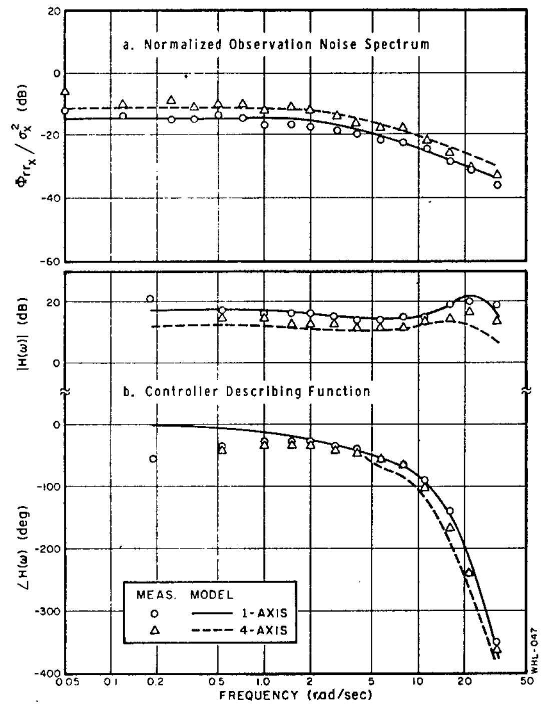


Figure 3. Effect of Attention Sharing on Frequency Response

Figure 2b shows good correspondence between model and experimental results. RMs error and control scores (not shown) were also generally well matched. Now, since the data were matched after the fact, we cannot, strictly speaking, consider these results as a demonstration of the model's predictive capability. Nevertheless, the ability to account quite well for the (rather considerable) changes in pilot describing functions and remnant spectra, with a fixed set of four independent model parameters, suggests that the model can be used a predictor of performance trends, at the very least. As described later, this capability of the model has been helpful in the design of simulator experiments.

As one might expect, the ability of the model to generate accurate predictions of pilot/vehicle performance, using the four parameter values shown at the bottom of Table 1, degrades as the complexity of the task increases. A recent study suggests that both the motor time constant and, to a lesser extent, observation noise ratio, increase with increasing order and/or delay in the plant dynamics [22]. This finding suggests that the current formulation of independent model parameters does not completely isolate pilot response capabilities from task parameters. Work is currently in progress to effect this isolation and thereby improve the absolute predictive capability of the model.

Attention-Sharing

A set of experiments was performed by Levison, Elkind, and Ward [4] to test the submodel for workload and attention-sharing described above. Subjects were provided with two 2-axis controls and four separated displays and were required to perform up to four concurrent linearly independent rate-control tracking tasks. The tasks were performed singly and in combination. When performing multiple tasks, fixation was maintained on a single display and the remaining three displays were tracked using peripheral vision (i.e., scanning was not allowed).

The single-variable tasks were used as "calibration" experiments to define numerical values for pilot-related parameters (including a "residual noise" term that was added to the expression for observation noise to account for peripheral viewing). With all model parameters held constant (except for noise/signal ratio), both attentional allocation and system performance were predicted for the 4-axis tracking task. Predictions were based on the assumptions that total task workload, f_t, remains constant for the single- and multi-axis tasks, and that the f_i sum to unity.

There was considerable interference among tasks; that is, subtask and total-task scores were greater when the tasks were performed concurrently than when performed singly. Model predictions for the multiple-task case were in good agreement with the data, especially for the total-task score (which is the quantity the subjects were instructed to minimize). Agreement with some of the subtask scores was less good, apparently because total-task performance was relatively insensitive to allocation of attention among the component tasks.

Figure 3 shows the effect of attention-sharing on the normalized observation noise (a linear transformation of the pilot's remnant spectrum) and on the pilot's describing function (i.e., control strategy). Results are presented for the foveally-fixated display. Model results agree quite well with the data and show that the effects of multiple-task requirements were to increase the observation noise spectrum (as predicted by the model for attention-sharing) and to decrease the amplitude ratio and increase high-frequency phase lag of the describing function (an adaptive response by the pilot to filter out some of the effects of increased observation noise.

The model relating attention to performance has also been validated for decision tasks and, to some extent, for combined decision and control tasks. Levison and Tanner [26] showed that the model for attention-sharing provided excellent agreement with the observed decrement in performance between a single decision task and two concurrent tasks. Predictions were less accurate for concurrent decision and control tasks; combined-task performance suggested that some interference was present, but to a lesser degree than predicted by the constraint of fixed total-task workload.

Multi-Sensory Cue Environments

The OCM has also been used to model continuous control performance in a multi-sensory cue environment. Levison and Junker [27,28] studied roll-axis tracking in disturbance-regulation and target-following tasks. They compared performance when only visual cues were available, to performance when the visual cues were augmented with platform motion cues. As shown in Figure 4, they found that the OCM could provide a task-independent framework for explaining performance under all experimental conditions. The availability of motion cues was modeled by augmenting the set of perceptual variables to include position, rate, acceleration, and acceleration-rate of the motion simulator. This straightforward informational treatment allowed the model to replicate the effects of motion cues on a variety of response measures, for both the target-following and disturbance-regulation tasks. Model results were obtained with a fixed set of independent model parameters.

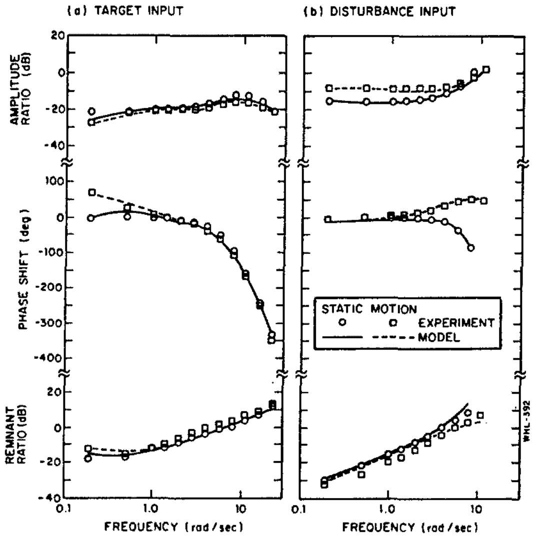


Figure 4. Comparison of Model and Experimental Frequency Response "Static": visual cues only. "Motion": combined visual and motion cues.

APPLICATIONS

The OCM has been applied -- mostly with regard to aircraft flight -- as a predictive and as a diagnostic tool. Areas of application include display design and evaluation, control design and evaluation, prediction of aircraft handling qualities, simulator design and evaluation, effects of environmental stress, and design of experiments. Examples in each area are cited below.

Display Design and Evaluation

The OCM has been used in numerous studies to evaluate the effectiveness of aircraft flight displays. In some cases, model predictions have been compared with data; in others, the OCM has been employed as a purely predictive tool.

Kleinman and Baron [29] analyzed a piloted approach-to-landing task to evaluate pictorial display requirements. This problem involved a time-varying information base for the pilot. The effects of different display formats and display symbology were predicted in cases where the aircraft was subjected to turbulence and/or constant updrafts. The ability of the pilot to estimate these external disturbances, and take the appropriate corrective action to minimize glide path errors was analyzed. Predictions of system performance were compared with data obtained in independent experimental investigations. The model-data agreements were excellent and demonstrated the model's ability to predict the time-varying adaptability of a pilot to updraft disturbances. In addition, the agreement between model results and data for cases in which there was no turbulence disturbing the aircraft, provided further evidence of the validity of the model for human randomness (remnant).

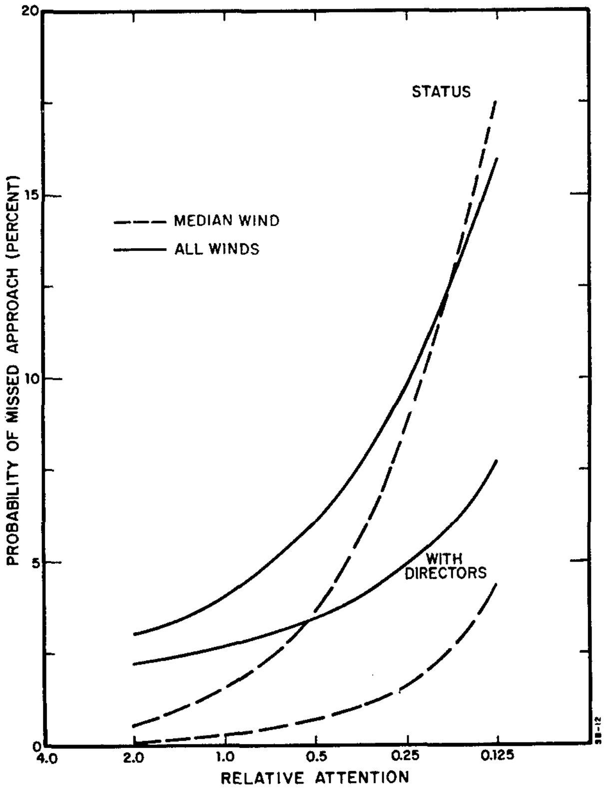


Figure 5: Effect of Lateral Director on Predicted Missed Approach Probability

Figure 5 illustrates a typical model application in a study by Baron and Levison [7] to develop an OCM-based display methodology for Short Takeoff and Landing aircraft. Two display configurations were considered: (1) "status", or standard symbolic display, and (2) a "flight director" display that combined bank angle, sideslip, and lateral path error in a single command display. Two conditions of atmospheric turbulence were also considered.

A quadratic cost criterion based on landing-approach requirements was formulated, and model predictions were obtained for each configuration as a function of observation noise/signal ratio. To generate the curves shown in the figure, noise/signal ratio was interpreted as relative attention per the submodel of Eq(4), and the quadratic performance index was transformed into a probability of missed approach (i.e., probability that certain system variables would be outside desired limits).

Figure 5 shows that the model predicted reasonable performance and workload trends; i.e., performance was expected to be superior with the director for a given level of attention (workload), and workload was expected to be less with the director for a given level of performance. (This study was analytic in nature and did not involve experimental verification).

A similar approach was followed to analyze both display and control characteristics for an aircraft with an advanced avionics configuration [6]; in this study, performance in a simulated windshear environment was explored both analytically and experimentally. Hess and Wheat used the OCM to analyze an electronic display for a helicopter [30].

The OCM has also been applied to the design of aircraft flight displays. Levison [31] suggested a procedure for designing flight directors in which the director laws would incorporate optimal processing of the display parameters as predicted by the OCM. Hess [32] has proposed a similar director design scheme which he has incorporated into a more formal display design procedure.

Control Design and Evaluation

Although display problems have received the most attention, other aspects of the system design problem have not been neglected completely. Levison and Houck [16] have explored the use of the model in analyzing control stick design problems in a vibration environment. Stengel and Broussard [33] have used the basic structure of the OCM along with some assumptions concerning sub-optimal adaptation to determine stability boundaries in high-g maneuvering flight. And, recently, Schmidt [34] has proposed a design procedure for stability augmentation systems based on closed-loop analysis with the OCM.

The study of Stengel and Broussard is particularly interesting because it illustrates an application of the model involving imperfect knowledge by the pilot of his vehicle response dynamics. The flight task explored in this study required the pilot to transition from a low angle-of-attack (AOA) flight phase to one involving high AOA. Manned simulation results were compatible with the notion that the test pilots retained internal models appropriate to low AOA flight (in which they had been trained) when in the high AOA phase.

Prediction of Pilot Opinion Ratings

Display and control design are but two of the factors that influence a vehicle's "handling qualities" as reflected in the pilot's ability to achieve acceptable system performance at reasonable levels of mental workload. Although aircraft handling qualities are specified, for the most part, in terms of vehicle response characteristics alone, the formal acquisition of subjective pilot opinion is an important aspect of aircraft elevation and acceptance. Thus, a need exists for a reliable analytic tool for predicting pilot opinion ratings, especially for new aircraft configurations and task environments.

Hess [35] noticed that, for a variety of experimental results that he matched with the OCM, the objective performance index varied monotonically with subjective pilot opinion ratings, and he suggested use of the OCM as a predictor for pilot ratings. Building on this effort, and on earlier work by Anderson [36] using servo-theory pilot models, Levison [37] suggested a scheme whereby model predictions of system performance and pilot workload could be combined to predict pilot opinion ratings.

Figure 6 illustrates application of the OCM-based prediction scheme. Performance/workload predictions are shown for eight of the vehicle configurations explored by Rickard [38] in a manned simulation study of longitudinal handling qualities for large transport aircraft in approach to landing. Model predictions generally mimicked the trends exhibited by subjective opinion ratings: the more favorably rated configurations tend to lie in the lower left-hand corner of the figure (corresponding to good performance and low workload), whereas the less favorably rated conditions lie toward the upper right. A set of "rating expressions" was derived to obtain, from the model results shown in Figure 6, "predictions" of opinion ratings that corresponded well with experimentally-obtained ratings.

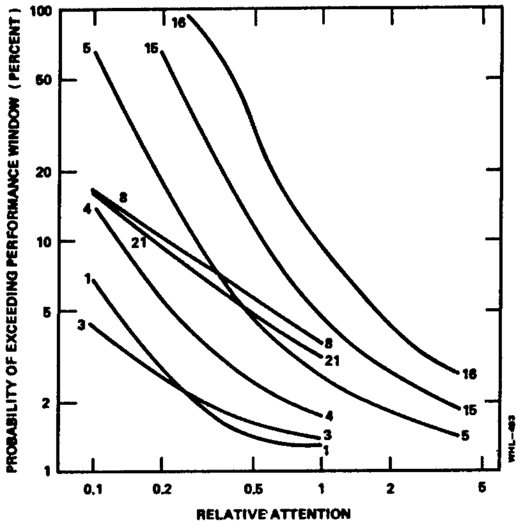


Figure 6. Prediction of Performance Vs. Attentional Workload for Eight Aircraft Configurations

A subsequent study was performed to provide a further test of the model-based handling qualities assessment scheme [39]. In this study, model predictions were obtained prior to a manned simulation study in which both objective as well as subjective performance measures were obtained. Although model predictions proved to be optimistic, compared to the experimental results, performance trends were reliably forecast. Specifically, the analytic scheme correctly identified the vehicle configurations that had the most adverse pilot opinion ratings and the largest rms excursions in aircraft state and control activity.

Simulator Design and Evaluation

The increased interest in flight simulators has spurred some additional extensions and applications of the model. Grunwald and Merhav [8] and Wewerinke [9,40] have incorporated mechanisms for describing the utilization of extra-cockpit visual scene cues in the OCM and have obtained preliminary experimental validation of their approaches. Although the subtleties and complexities associated with human perception of a complex scene are by no means resolved, these studies do suggest that the OCM could be useful for analyzing closed-loop control behavior based on visual scene cues.

In a somewhat different vein, Baron, Muralidharan and Kleinman [41] used the OCM to develop a closed-loop model for analyzing engineering requirements for flight simulators. They predicted the effects on performance of certain simulation design parameters, such as integration scheme, sample rate, etc. Model predictions were later verified in an empirical study by Ashworth [42].

Environmental Stress

Military operational environments may subject the pilot or gun operator to substantial physical stress. In some cases, the stress is a direct consequence of the flight task (e.g., vibration, sustained high acceleration); in other cases, stress may be induced by an opponent as a defensive measure (e.g., optical countermeasure). Such considerations have motivated application of the OCM to tasks involving actual and simulated environmental stress.

A series of studies was conducted to develop a methodology for modeling the effects of high-frequency vibration on pilot response behavior and total system performance [16,43-45]. This effort led to a model structure which combines the OCM with a biodynamic model of the human operator [46]. As part of this structure, a set of rules were developed for relating certain OCM parameters (specifically, time delay and motor noise covariance) to biodynamic response parameters.

Korn, Ephrath, and Kleinman [47] have recently applied the OCM to a study of the effects of sustained positive linear acceleration (G_z stress. The acceleration stress degraded system performance; these effects were modeled by increases in all pilot-related OCM parameters. In a subsequent study, Korn and Kleinman [48] modeled the effects of lateral acceleration stress (G_y) by changes in OCM parameters consistent with the notion that the subjects were willing to tolerate larger errors when subjected to acceleration stress.

Zacharias and Levison [5] have explored use of the OCM as a diagnostic tool for identifying the particular type of information-processing decrements caused by environmental stress. They showed that their identification procedure would be adequate to differentiate between an increase in information-processing time and a decrease in visual resolution capabilities.

Design of Simulation Experiments

This author and his colleagues have routinely used the OCM in the design phase of simulation experiments. Typically, the model is used to select certain experimental parameters to maximize the effect of the principal experimental variables.

A design application of this type has been documented by Junker and Levison [49], who used the OCM to design the motion-cue study described above in the discussion of model validation. Since one of the goals of that study was to determine how to model the effects of motion cues, pre-experiment model runs were made to determine forcing-function spectra and performance objectives so that (1) motion would significantly improve performance in one experimental condition, and (2) motion cues would have no significant effect in another. Although pre-experiment model results did not match experimental results as closely as shown in Figure 4, the trends of the results were quite well predicted, and a useful experiment was consequently performed.

The OCM was used in a subsequent study to design an experiment exploring the pilot's use of the "tilt cue" (i.e., the apparent rotation of the gravity vector that is felt when a ground-based simulator undergoes pure roll-axis displacement) [50]. Considerable pre-experiment analysis was required to find a set of vehicle dynamics such that the tilt cue would significantly influence tracking performance. Again, experimental performance trends were as predicted by the model.

FURTHER DEVELOPMENT AND APPLICATION

Studies are currently in progress to extend the capabilities of the optimal control model and to apply it to a wider variety of tasks. One such study is expected to indicate how the predictive accuracy of the OCM may be improved (Air Force Contract No. F33615-81-C-0517). As noted in the discussion on model validation, reliable performance trends can often be obtained using a fixed set of values for pilot-related model parameters. Nevertheless, if we consider the full range of laboratory results, certain systematic deviations in these parameters are observed.

The consistency of such parameter deviations suggests, first, that parameters identified as "pilot-related" reflect, in part, the interaction between the human operator and the task environment; and, second, that these deviations should be predictable with a suitable model structure. Therefore, in order to improve the predictive capabilities of the OCM -- especially in terms of predicting absolute performance -- attempts will be made to isolate parameters that more nearly reflect human response limitations.

The study referenced above will also address the development of models for "learning behavior" (i.e., control-strategy development) -- one of the important remaining theoretical frontiers in manual control. Models of this sort have ready application to the design of training simulators: e.g., predicting for a given type of flight task whether or not whole-body motion cues will reduce flight training time. It is not necessarily sufficient to predict the effect of motion cues on the performance of a fully trained pilot; rather, the question is the extent to which motion cues might allow the trainee more rapidly to develop the correct response strategy (say, by more rapidly developing a correct "internal model" of the task environment).

One approach being considered is to include a fourth adaptive element to the pilot model: an optimal "identifier" to account for development of the pilot's internal model in a given task situation. This model element would account both for the rate of learning as well as the asymptotic structure of the internal model (which, in complex tasks, may be different from the true system).

Because of high lateral accelerations that may be induced by modern military aircraft that allow direct control of side force, techniques are desired to optimize the design of control sticks in high-g environments. To this end, a study is in progress to better define the pilot/stick interface (Air Force Aerospace Medical Research Laboratory Project No. L72311708). Three control devices, ranging from a near-isometric device to one that has only weakly restrained displacement, are to be explored in a fixed-base simulation study. Three electrical gains will be explored for each stick. The results of this study are expected to yield, in part, a better definition of the motor sub-model of the OCM.

A study is also in progress to better define the perceptual limitations of the human operator (Air Force Contract No. F33615-81-C-0515). This analytical and experimental study will explore the pilots of non-visual motion cues (platform motion and g-seat cuing) as well as the use of visual scene cues relevant to the task of high-speed contour-following. Preliminary work in the area of modeling visual scene cues has been reported [10].

The major opportunity for extending the OCM to other tasks is in the area of supervisory control. These control problems, as we have noted, involve monitoring, detection, decision-making, and discrete and/or infrequent control. Most often, the systems are highly automated, require more than one operator and are extremely complex. The principal feature of the OCM that is useful for these application is its information processing structure, but the underlying, normative modelling perspective is also important. The OCM has been applied to tasks of this sort [21,26,51-54] and additional studies are contemplated to develop and apply an OCM-based model to the study of flight crew performance.

The application of the OCM, or its derivatives to such problems, is in its early stages and many questions remain to be answered. Among the more important are questions concerning the human's internal model for such large scale systems, the appropriate control and decision cost functionals for these problems, the modeling of attention-sharing strategies in time-varying situations, and the appropriate level for incorporating and modeling aspects of the tasks that are important but are not likely to be treated using the same techniques (e.g., certain types of intellectual activity or procedural tasks). Finally, there is a need for additional data and, indeed, for appropriate experimental paradigms, for establishing these models. It must be recognized that validation of such complex models to the degree that manual control models have been and can be validated, will not be possible for some time, for both theoretical and practical reasons. One may not expect, therefore, that supervisory control models that are "predictive", in the same sense as the OCM, will be developed in the near future. Nonetheless, it should be possible to develop models of supervisory control that will capture many essential features of tasks of interest and will prove to be useful design, analysis, and evaluation tools.

REFERENCES

1. Kleinman, D.L., S. Baron and W.H. Levison, "An Optimal-Control Model of Human Response, Part I: Theory and Validation", Automatica, Vol. 6, pp. 357-369, 1970.

2. Kleinman, D.L., S. Baron and W.H. Levison, "A Control Theoretic Approach to Manned-Vehicle Systems Analysis", IEEE Trans. on Auto. Control, Vol. AC-16, pp. 824-833, No. 6, December 1971.

3. Levison, W.H., "The Effects of Display Gain and Signal Bandwidth on Human Controller Remnant", AMRL-TR-70-93, Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, March 1971.

4. Levison, W.H., J.I. Elkind and J.L. Ward, "Studies of Multivariable Manual Control Systems: A Model for Task Interference", NASA CR-1746, May 1971.

5. Zacharias, G.L. and W.H. Levison, "A Performance Analyzer for Identifying Changes in Human Operator Tracking Strategies", AMRL-TR-79-17, Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, March 1979.

6. Levison, W.H., "Analysis and In-Simulator Evaluation of Display and Control Concepts for a Terminal Configured Vehicle in Final Approach in a Windshear Environment", NASA CR-3034, August 1978.

7. Baron, S. and W.H. Levison, "A Manual Control Theory Analysis of Vertical Situation Displays for STOL Aircraft", NASA CR-114620, April 1973.

8. Grunwald, A.J. and Merhav, S.J., "Vehicular Control by Visual Field Cues--Analytical Model and Experimental Validation", IEEE Trans. Systems, Man and Cybernetics, Vol. SMC-6, No. 12, March 1976, pp. 835-845.

9. Wewerinke, P.H., "A Theoretical and Experimental Analysis of the Outside World Perception Process", Proc. of Fourteenth Annual Conference on Manual Control, NASA Conference Publication 2060, November 1973, pp. 535-556.

10. Zacharias, G.L. and W.H. Levison, "A Model-Based Procedure for Determining Visual Cue Requirements", Proc. of the 1981 Image Generation/Display Conference II, Air Force Human Resources Laboratory, June 1981.

11. Baron, S. and W.H. Levison, "An Optimal Control Methodology for Analyzing the Effects of Display Parameters on Performance and Workload in Manual Flight Control", IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-5, No. 4, pp. 423-430, July 1975.

12. Baron, S. and W.H. Levison, "Display Analysis with the Optimal Control Model of the Human Operator", Human Factors, Vol. 19, No. 5, pp. 437-457, October 1977.

13. Curry, R.E., W.C. Hoffman, and L.R. Young, "Pilot Modeling for Manned Simulation", AFFDL-TR-76-124, Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio, 1976.

14. Zacharias, G.L., "Motion Cue Models for Pilot-Vehicle Analysis", AMRL-TR-78-2, Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, May 1978.

15. Levison, W.H., S. Baron and A.M. Junker, "Modeling the Effects of Environmental Factors on Human Control and Information Processing", AMRL-TR-76-74, Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, August 1976.

16. Levison, W.H. and P.D. Houck, "Guide for the Design of Control Sticks in Vibration Environments", AMRL-TR-74-127, Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, February 1975.

17. Jex, H.R. and R.E. Magdaleno, "Biomechanical Models for Vibration Feed-Through to Hands and Head for a Semisupine Pilot", Aviation Space and Environmental Medicine, Vol. 49, No. 1, pp. 304-317, January 1978.

18. Levison, W.H., "A Model for Mental Workload in Tasks Requiring Continuous Information Processing", Mental Workload Its Theory and Measurement, Plenum Press, New York and London, 1979.

19. Levison, W.H. and D.L. Kleinman, "A Model for Human Controller Remnant", Proc. of the Fourth Annual Conference on Manual Control, University of Michigan, Ann Arbor, MI, March 1968, pp. 3-15.

20. Wewerinke, P.H., "Human Operator Workload for Various Control Situations", Proc. of Tenth Annual Conference on Manual Control, Wright-Patterson Air Force Base, OH, 1974.

21. Kleinman, D.L. and R.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.

22. Levison, W.H., "Effects of Whole-Body Motion Simulation on Flight Skill Development", Report No. 4645, Bolt Beranek and Newman, Inc., Cambridge, MA, November 1981.

23. Levison, W.H., Zacharias, G.L., Porterfield, J., Monk, D., and Arbak, C., "A Comparison of Head and Manual Control for a Position-Control Pursuit Tracking Task", Proc. of the Seventeenth Annual Conf. on Manual Control, Los Angeles, CA, June 1981.

24. Levison, W.H., "Modeling the Pilot's Use of Motion Cues During Transient Aircraft Maneuvers", BBN Report No. 4312, Bolt Beranek and Newman Inc., Cambridge, MA, March 1980.

25. Levison, W.H., "Methods for Identifying Pilot Dynamics", Proc. of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, Edwards Air Force Base, CA, January 1982.

26. Levison, W.H. and Tanner, R.B., "A Control-Theory Model for Human Decision Making", NASA CR-1953, December 1971.

27. Levison, W.H. and Junker, A.M., "A Model for the Pilot's Use of Motion Cues in Steady-State Roll-Axis Tracking Tasks", presented at the AIAA Flight Simulation Technologies Conference, Arlington, TX, September 1978.

28. Levison, W.H. and Junker, A.M., "A Model for the Pilot's Use of Motion Cues in Roll-Axis Tracking Tasks", AMRL-TR-77-40, Aerospace Medical Research Laboratory, June 1977.

29. Kleinman, D.L. and Baron, S., "Analytic Evaluation of Display Requirements for Approach to Landing", NASA CR-1952, November 1971.

30. Hess, R.A. and Wheat, I.W., "A Model-Based Analysis of a Display for Helicopter Landing Approach", IEEE Trans. on Systems, Man and Cybernetics, pp. 505-511, Vol. SMC-6, No. 7, July 1976.

31. Levison, W.H., "A Model-Based Technique for the Design of Flight Directors", Proc. of the Ninth Annual Conference on Manual Control, May 1973.

32. Hess, R.A., "Analytical Display Design for Flight Tasks Conducted Under Instrument Meteorological Conditions", IEEE Trans. on Systems, Man and Cybernetics, SMC-7, No. 6, June 1977, pp. 453-461.

33. Stengel, R.F. and Broussard, J.R., "Prediction of Pilot-Aircraft Stability Boundaries and Performance Contours", IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-8, May 1978, pp. 349-356.

34. Schmidt, D.K., "Optimal Flight Control Synthesis via Pilot Modeling", Journal of Guidance and Control, Vol. 2, pp. 308-312, July-August 1979.

35. Hess, R.A., "Prediction of Pilot Opinion Ratings Using an Optimal Pilot Model", Human Factors, 1977, 19(5), 459-475.

36. Anderson, R.O., "A New Approach to the Specification and Evaluation of Flying Qualities", Wright-Patterson Air Force Base, AFFDL-TR-69-120, 1970.

37. Levison, W.H., "A Model-Based Technique for Predicting Pilot Opinion Ratings for Large Commercial Transports", Report No. 4153, Bolt Beranek and Newman Inc., Cambridge, MA, July 1979.

38. Rickard, W.W., "Longitudinal Flying Qualities in the Landing Approach", Proc. of the Twelfth Annual Conference on Manual Control, NASA TMX-73, 170, May 1976.

39. Levison, W.H. and Rickard, W.W., "Closed-Loop Criteria for Assessing Longitudinal-Axis Handling Qualities of Transports in Final Approach", Proc. of the AIAA Guidance and Control Conf., Albuquerque, NM, August 19-21, 1981.

40. Wewerinke, P.H., "The Effect of Visual Information on the Manual Approach and Landing", Proc. of the Sixteenth Annual Conf. on Manual Control, MIT, May 5-8, 1980.

41. Baron, S., Muralidharan, R. and Kleinman, D.L., "Closed-Loop Models for Analyzing Engineering Requirements for Simulators", BBN Report No. 3718, Bolt Beranek and Newman Inc., Cambridge, MA, May 1978.

42. Ashworth, B.G., McKissick, B.T. and Parrish, R.V., "The Effects of Simulation Fidelity on Air-to-Air Tracking", Proc. of the Fifteenth Annual Conf. on Manual Control, WPAFB, Ohio, March 1979.

43. Levison, W.H., "Model for Human Controller Performance in Vibration Environments", Aviation Space and Environmental Medicine, Vol. 49, No. 1, pp. 321-328, January 1978.

44. Levison, W.H. and Harrah, C.B., "Biochemical and Performance Response of Man in Six Different Directional Axis Vibration Environments", AMRL-TR-77-71, Aerospace Medical Research Laboratory, WPAFB, September 1977.

45. Levison, W.H., "Biomechanical Response and Manual Tracking Performance in Sinusoidal, Sum-of-Sines, and Random Vibration Environments", AMRL-TR-75-94, Aerospace Medical Research Laboratory, WPAFB, Ohio, April 1976.

46. Berliner, J.E. and W.H. Levison, "PIVIB: A Computer Program for Analysis of Pilot Biodynamic and Tracking Response to Vibration", AMRL-TR-77-72, Aerospace Medical Research Laboratory, WPAFB, Ohio.

47. Korn, J., Ephrath, A.R. and Kleinman, D.L., "Human Operator Modeling for Aerial Tracking", Final Technical Report TR-79-16, Aeromedical Research Laboratory, Wright-Patterson Air Force Base, Ohio, November 1979.

48. Korn, J. and Kleinman, D.L., "Modeling Lateral Acceleration Effects on Pilot Performance", Proc. of the Sixteenth Annual Conference on Manual Control, MIT, May 5-8, 1980.

49. Junker, A.M. and Levison, W.H., "Recent Advances in Modelling the Effects of Roll Motion on the Human Operator", Aviation Space and Environmental Medicine, Vol. 49, No. 1, pp. 328-335, January 1978.

50. Levison, W.H. and Junker, A.M., "Use of a Tilt Cue in a Simulated Heading Tracking Task", Proc. of the Fourteenth Annual Conference on Manual Control, Los Angeles, CA and Ames Research Center, Moffett Field, CA, November 1978.

51. Phatak, A. and Kleinman, D.L., "Current Status of Models for the Human Operator as a Controller and Decision Maker in Manned Aerospace Systems", Automation in Manned Aerospace Systems, AGARD Proc. No. 114, Dayton, Ohio, October 1974.

52. Gai, E.G. and Curry, R.E., "Failure Detection by Pilots During Automatic Landing; Models and Experiments", Proc. of the Eleventh Annual Conf. on Manual Control, NASA TMX-62, 464, May 1975, pp. 78-93.

53. Gai, E.G. and Curry, R.E., "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. 85-93.

54. Muralidharan, R. and Baron, S., "Combined Monitoring, Decision and Control Model for the Human Operator in a Command and Control Task", Proc. of the Fourteenth Annual Conf. on Manual Control, NASA Publication No. 2060, November 1978, pp. 647-662.

TIME DOMAIN IDENTIFICATION OF AN OPTIMAL CONTROL PILOT MODEL WITH EMPHASIS ON THE OBJECTIVE FUNCTION

DAVID K. SCHMIDT
SCHOOL OF AERONAUTICS AND ASTRONAUTICS
PURDUE UNIVERSITY

PRESENTED AT THE
AFFTC/NASA DRYDEN/AIAA WORKSHOP
ON FLIGHT TESTING TO IDENTIFY PILOT
WORKLOAD AND PILOT DYNAMICS
EDWARDS AFB, CA 93523
JANUARY 19-21, 1982

Introduction

In this paper, we will propose a method for the identification of the pilot's control compensation using time domain techniques. From this information we hope to infer a quadratic cost function, supported by the data, that represents a reasonable expression for the pilot's control objective in the task being performed, or an inferred piloting "strategy". (Note here that we are using the term strategy as synonomous with control objective, and not with control law.)

The ultimate goals of this research topic include a better understanding of the fundamental piloting techniques in complex tasks, such as landing approach; the development of a metric measurable in simulations and flight test that correlate with subjective pilot opinion; and to further validate pilot models and pilot-vehicle analysis methods. At this time we will present the methodology and some preliminary numerical results.

The Pilot Model and Objective Function

The analyses relies on the well-known [1] optimal-control theoretic technique for modeling the human pilot's manual control function. The hypothesis upon which it is based is that the well trained, well motivated pilot chooses his control inputs (e.g. stick force) to meet the pilot's (internal) objective in the task, subject to his human limitations. His objective is further assumed to be expressible in terms of a quadratic "cost" function

$$J_p = E\left\{\lim_{T\to\infty} \frac{1}{T}\int_0^T \left(Y_p^T QY_p + u_p^T Ru_p + \dot{u}_p^T G\dot{u}_p\right) dt\right\} \qquad (0)$$

where Y_p = vector of pilot's observed variables (e.g., attitude, acceleration)

u_p = vector of pilot's control inputs

Q,R,G = Pilot-Selected (internal) weightings

The human limitations modeled include information-acquisition and processing time delay, observation and control input errors, and neuromuscular dynamics. A block diagram of the resulting model structure is shown in Figure 1.

The components of this model may be grouped into two parts, one dealing with the information acquisition and state estimation, and one related to the control law or control policy operating on the estimated state. As has been shown in the references on this modeling approach, the "solution" for the pilot's control inputs, as predicted by the model, is expressed as

$$\dot{u}_p = g_x^T \hat{x} + g_u^T u_p + \nu_u$$

where $\hat{x}$ = internal estimate of the system state

g_x, g_u = control gains

ν_u = motor noise, or control input errors

(Readers unfamiliar with the further details of the model are referred to the reference.)

The key points germain to this analysis are that the above equation is a mathematical expression representing the pilot's overt control actions (u_p), and these control actions are measurable experimentally. Furthermore, the gains g_x and g_u are functions of the plant (vehicle) dynamics and his objective function, and thereby represent his control "techniques", level of skill, and familiarity with the vehicle dynamics.

Another factor of importance is that not only is the objective function, from which the gains are determined, a mathematical part of a pilot control model, but it's resulting magnitude obtained from exercising the <u>model</u> has been found to correlate with the subjective pilot opinion obtained from <u>simulation</u> and <u>flight test</u>. Such a correlation is shown in Figure 2, as an example, taken from Refs. 2 and 3. This of course assumes one has been able

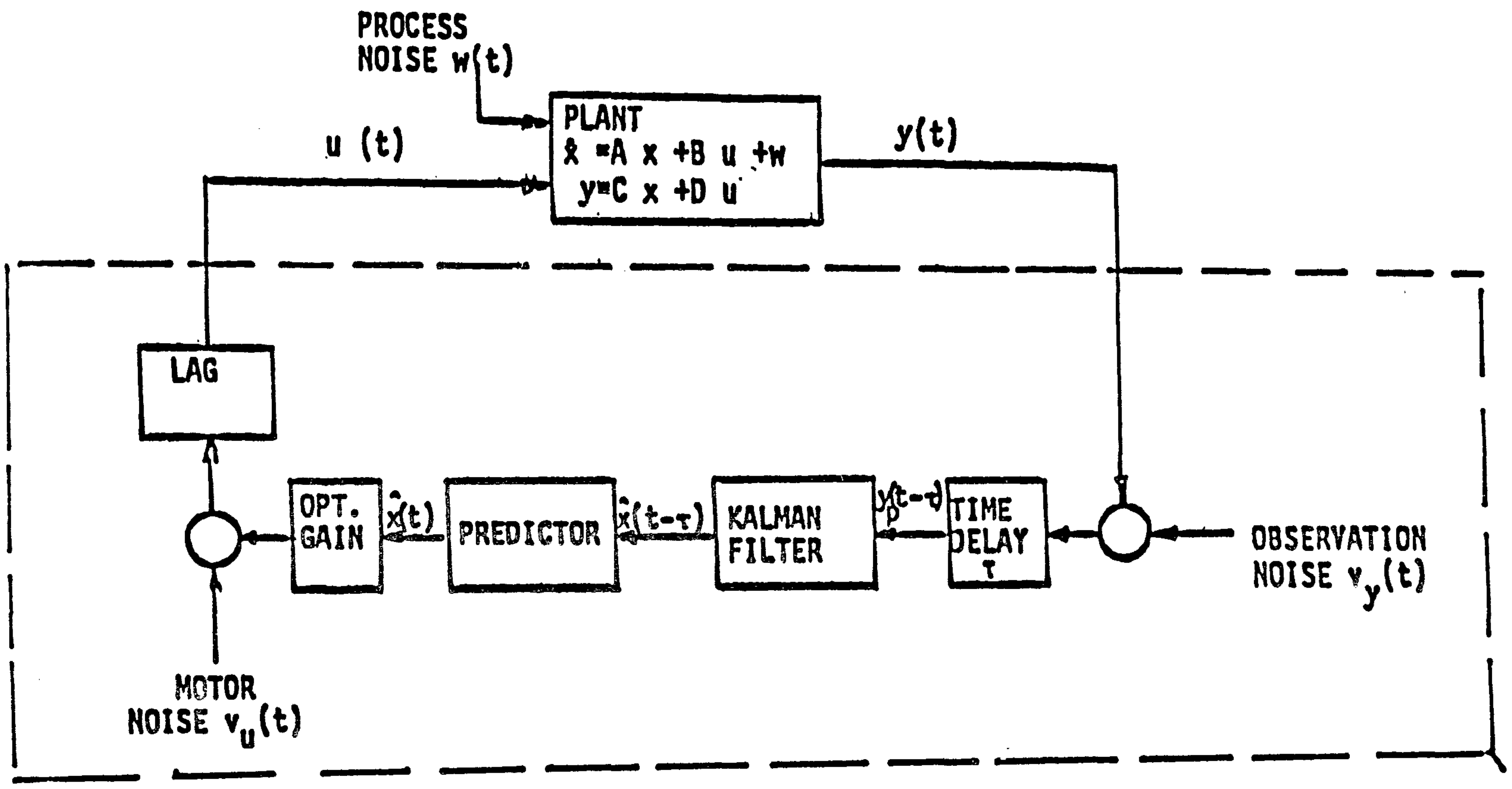


Figure 1, Model Structure

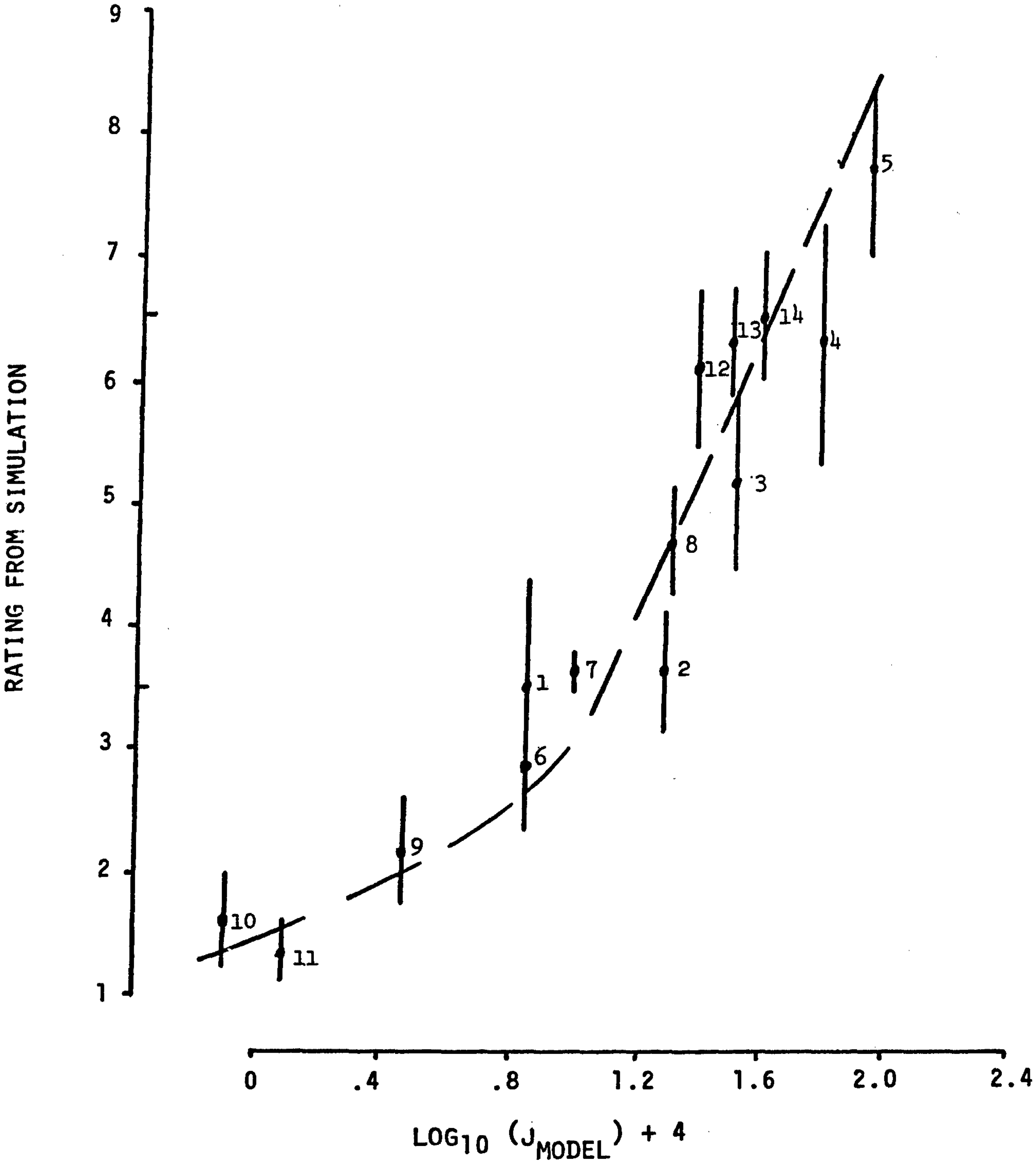


Figure 2, Rating Correlation

to correctly express the pilot's (internal) cost function, which is in fact his strategy that depends on his perception of the task. Now this is easy to do in simple laboratory tasks in which the subject has been instructed to minimize some displayed error, for example. But it is not at all clear just what flight parameters are being "regulated" or "tracked", other than ILS glide slope and localizer error in the case of landing approach. This is but one example, other complex piloting tasks might be considered equally as well.

The Identification Procedure

We seek then a method by which we may identify those pilot parameters that reflect his control techniques, or control strategy. Referring back to the pilot model control law, or

$$\dot{u}_p = g_x^T \hat{x} + g_u^T u_p + \nu_u$$

we note that the gains g_x operate on the estimated state $\hat{x}$. Now the separation principle of optimal estimation and control theory states that the control gains (g_x, g_u) are independent of the state estimation process. Further, the optimal state estimator, in general and in the pilot model, is independent of the overall objective function being minimized by the controller (estimator and control) law. Therefore, if we are mainly after the pilot's control strategy as expressed by, or at least a function of, his objective function, we need only to focus on the gains (g_x, g_u) and not on those variables related only to the state estimator. These latter variables include the time delay, and observation and motor noise covariance matrices, parameters of interest in the identification technique of Levison [4], for example. If our approach is successful, fewer parameters must be identified from the data, which is always an advantage, but the parameters affecting the estimation process are assumed.

The identification method proposed is as follows. The control law expressed previously, may be rewritten as

$$\dot{u}_p = g_x^T x - g_x^T \varepsilon + g_u^T u_p + \nu_u$$

where ε = error in estimating the true (actual) state x. Note that along with the pilot's control u_p, these true states, such as angle of attack or pitch attitude are measureable, but the state estimate, $\hat{x}$, is a quantity internal in to pilot, as modeled. Hence $\hat{x}$ is not measurable ---nor are ε or ν_u. Transposing the above, multiplying by $x = \text{col}\left[x, u_p\right]$, and taking expected values yields

$$\left[\begin{array}{c} E(x\dot{u}_p^T) \\ \hdashline E(u_p\dot{u}_p^T) \end{array}\right] = \left\{ \left[\begin{array}{c:c} E(xx^T) & E(xu_p^T) \\ \hdashline E(u_p x^T) & E(u_p u_p^T) \end{array}\right] - \left[\begin{array}{c:c} E(x\varepsilon^T) & 0 \\ \hdashline E(u_p \varepsilon^T) & 0 \end{array}\right] \right\} \left[\begin{array}{c} g_x \\ \hdashline g_u \end{array}\right]$$

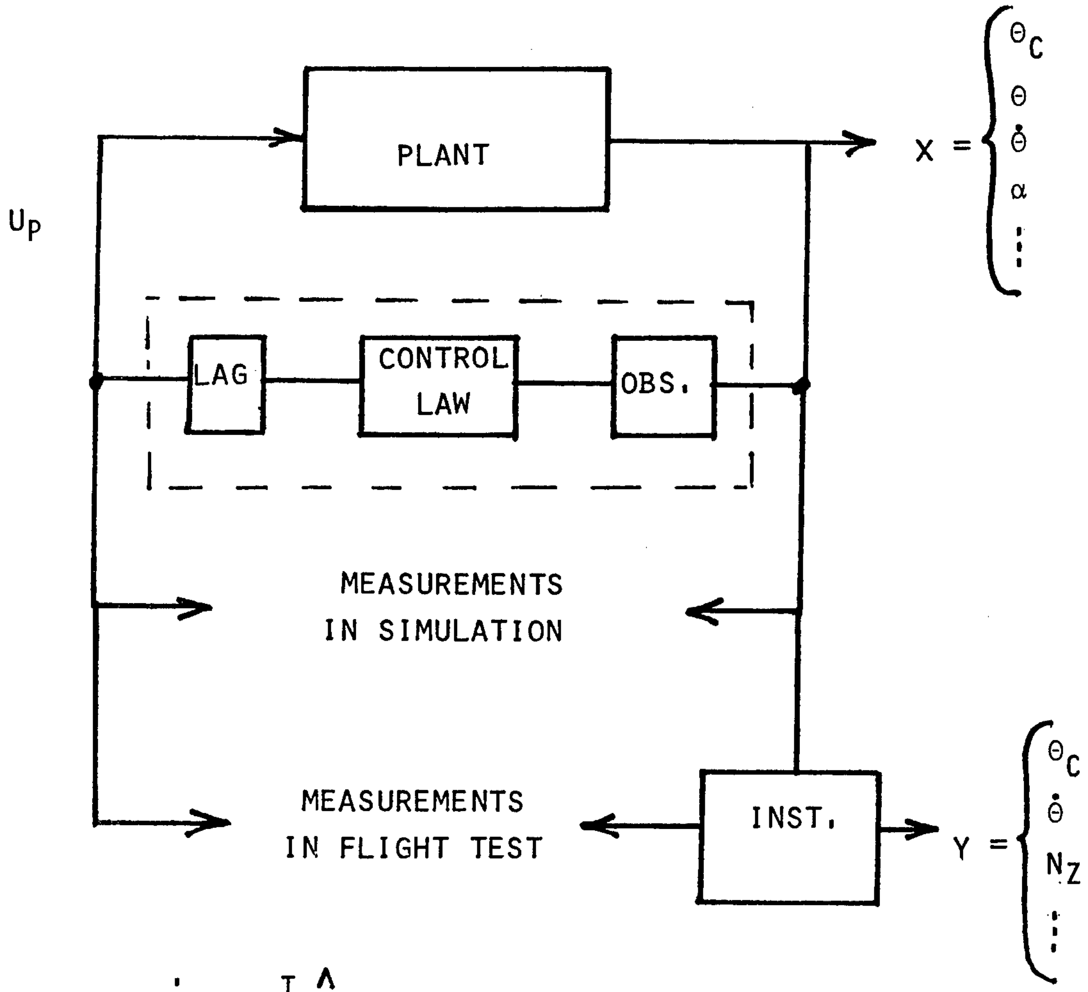
$$+ \left[\begin{array}{c} E(x\nu_u^T) \\ \hdashline E(u_p \nu_u^T) \end{array}\right]$$

$$\text{or } N_{\dot{u}} = M \left[\begin{array}{c} g_x \\ \hdashline g_u \end{array}\right] + N_{\nu_u} \tag{I}$$

Now to evaluate these matrices we note first that, in a simulation at least, the vectors x(t) and $u_p(t)$ are measurable, so estimates of their covariance matrices (e.g., $E(xx^T)$) may be obtained from measurements of sampled time histories. (Also, in this paper we assume that good estimates of $\dot{u}_p$ are available from filtered measurements of u_p. The details of accomplishing this filtering are under current investigation, but digital techniques as well as analog methods are still available.) For reference, refer to Figure 3.

EXPERIMENTAL CONSIDERATIONS

TRACKING TASK - ATTITUDE, ACCEL. ...



$$\dot{U}_P = G_X^T \hat{X}_P + G_U U_P + V_U$$

Figure 3

With regard to the remaining terms involving ε and ν_u, both are not measurable and need attention. To resolve this consider the complete system dynamics model by the relation

$$\dot{x} = Ax + Bu_p + w$$

and

$$\dot{u}_p = g_x^T x - g_x^T \varepsilon + g_u^T u_p + \nu_u$$

where the relation between state and estimate, or $\hat{x} = x - \varepsilon$ has been employed. The pilot's internal state estimation error, ε, is treated as follows. Define $\varepsilon_u = u_p - \hat{u}_p$ to be the error in estimation of the pilot's own control input, and then let

$$\bar{\varepsilon} = \text{col}\left[\varepsilon, \varepsilon_u\right]$$

Now the covariance of $\bar{\varepsilon}$ may be shown to be governed by the relation

$$\text{cov}\ (\bar{\varepsilon}) = E(\bar{\varepsilon}\,\bar{\varepsilon}^T) \triangleq P$$

$$\dot{P} = A_1\ P + PA_1^T + W_1$$

Also we have

$$A_1\ \Sigma + \Sigma\ A_1^T + W_1 - \Sigma C^T\ V_y^{-1}\ C\Sigma = 0;\ \Sigma = \text{cov}\ (e_{KF})$$

and

$$A_1 = \left[\begin{array}{c|c} A & B \\ \hline 0 & g_u \end{array}\right]$$

C = pilot's observation matrix

$$Y_p = C\begin{bmatrix} x(t-\tau) \\ u(t-\tau) \end{bmatrix} + \nu_y$$

These relations are all obtained from Ref. (5) and from the pilot model equations given in Ref. (1). Here e_{KF} is the Kalman filter estimation error for the delayed state, Σ the covariance of e_{KF}, and

$$W_1 = \left[\begin{array}{c|c} W & 0 \\ \hline & V_u \end{array}\right]$$

Also W is the covariance of the plant disturbance w, and V_u and V_y are motor noise and measurement noise covariance, respectively, all assumed known. Now the $\dot{P}$ equation may be integrated over the time delay τ, with the initial condition on P from $P(0) = \Sigma$, the Kalman filter error covariance. Now, since the predictor has the property that $E(\hat{x}\ \bar{\varepsilon}^T) = 0$, we have

$$E\left\{\left[\begin{array}{c} x \\ \hline u_p \end{array}\right] \bar{\varepsilon}^T\right\} = E(\bar{\varepsilon}\ \bar{\varepsilon}^T) = P$$

So then the terms $E(x\ \varepsilon^T)$ and $E(u_p\varepsilon^T)$ are available from P, and these are required to form M.

Finally, it can be shown (Ref. (5)), pg. 331) that with the processes w and ν_u uncorrelated we have in this case

$$E(x\nu_u^T) = 0$$

$$E(u_p\nu_u^T) = \frac{1}{2} V_u$$

Returning then to the estimation of the gains (equation I), we see that all the terms in the matrices $N_{\dot{u}}$, N_{ν_u} and M may be calculated, either analytically or from the measurements of x, u_p (and $\dot{u}_p$). The estimate for the gain vector is then

$$\left[\begin{array}{c} g_x \\ \hline g_u \end{array}\right]_{est} = M^{-1}\left[N_{\dot{u}} - N_{\nu_u}\right] \qquad \text{(II)}$$

Note finally that the matrix M is formed from two matrices

$$M = M_x - M_{cor}$$

where the M_{cor} and N_{ν_u} matrices may be thought of as corrections added to a basic least-squares technique. The potential importance of these terms (M_{cor} and N_{ν_u}) will be demonstrated in an example later.

The algorithm is as follows:

1) Select noise covariance matrices, W, V_u, and V_y

2) Select a time delay τ, neuromuscular time constant τ_N (or matrix $T_N = g_u^{-1}$).

3) Form A_1 and solve for Kalman Filter error covariance Σ.

4) Solve for covariance matrix $P(\tau)$ and then the $E(\bar{\varepsilon}\,\bar{\varepsilon}^T)$ is available. (Note, all these steps may be accomplished before or after the experimental data is obtained.)

5) Perform experiment to obtain state and control (and control rate) time histories.

6) From the time histories, obtain estimates for $E(xx^T)$, $E(xu_p^T)$, $E(u_p u_p^T)$, $E(x\dot{u}_p^T)$ and $E(u_p\dot{u}_p^T)$, or the matrices M_x and $N_{\dot{u}}$.

7) Identify M_{cor} and N_{ν_u} in $E(\bar{\varepsilon}\,\bar{\varepsilon}^T)$ found in step 4.

8) Form $M = M_x \div M_{cor}$ and determing $\left[\frac{g_x}{g_u}\right]_{est}$ from Equation II.

9) Check g_u vs T_N^{-1} (selected in 2 above) and iterate (steps 2-8) again as necessary. Note now that selecting τ_N affects the solution for Σ and $P(\tau)$, along with the effective V_u or

$$V_{u_{eff}} = T_N^{-1}\, V_u \left(T_N^{-1}\right)^T$$

while selecting τ only affects $P(\tau)$ in the procedure.

Comparison to Classical Results

It is interesting to note that the "corrections" performed by including M_{cor} and N_{ν_u} are qualitatively related to an identification technique (discussed in Ref. 6, and elsewhere) used to determine the human describing function in a compensatory task, which goes back to the development of the "crossover model" of McGruer et al. Shown in Figure 4 is a schematic of this situation, showing the closed-loop tracking of some commanded θ_c. Measurements may be taken of $\theta_c(t)$, $\varepsilon(t)$, $u_p(t)$, and $\theta(t)$ and manipulated in the frequency domain to obtain frequency spectra

$$G_1(j\omega) = U_p(j\omega)/\theta_c(j\omega)$$
$$G_2(j\omega) = \varepsilon(j\omega)/\theta_c(j\omega)$$
$$G_3(j\omega) = U_p(j\omega)/\varepsilon(j\omega)$$

Now, in this model the pilot's control is considered to consist of two parts, one correlated with the input θ_c, the other uncorrelated with the input. The latter component was defined to be "remnant." Mathematically,

$$u_p(j\omega) = Y_p(j\omega)\ \varepsilon(j\omega) + r(j\omega) \text{ and } r(j\omega)/\theta_c(j\omega) \to 0 \text{ in effect.}$$

Block diagram manipulation leads then to the desired relation

$$Y_p(j\omega) = G_1(j\omega)/G_2(j\omega)$$

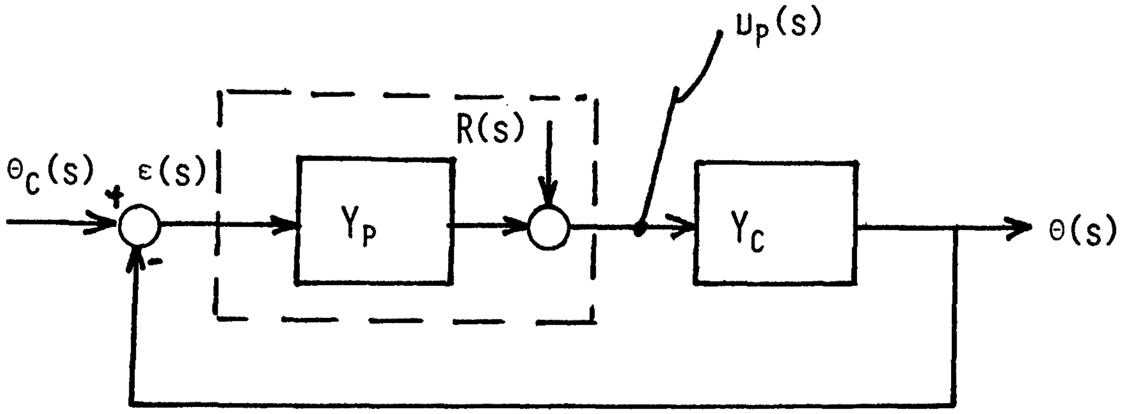
rather than the simpler, and incorrect, expression $Y_p(j\omega) = G_3(j\omega)$. This was due to the presence of remnant $r(j\omega)$ in the <u>measured</u> control input, and the necessity to eliminate it's effect by defining it as the uncorrelated component of u_p, and using this property. Comparing to our control law, transformed just for discussion purposes, we have

$$\dot{u}_p(j\omega) = \underbrace{g_x^T\ x(j\omega) + g_u^T\ u_p(j\omega)} - \underbrace{g_x^T\ \bar{\varepsilon}(j\omega) + \nu_u(j\omega)}_{\substack{\text{unmeasurable}\\ \text{separately}}}$$

compared to

$$u_p(j\omega) = Y_p(j\omega)\ \varepsilon(j\omega) + \underbrace{r(j\omega)}_{\substack{\text{unmeasurable}\\ \text{separately}}}$$

CLASSICAL RESULTS



MEASURE $U_P(s) = Y_P\ \varepsilon(s) + R(s)$

$$Y_P = G_1(JW)/G_2(JW) \neq G_3(JW)$$

WHERE $G_1(s) = U_P(s)/\Theta_C(s)$

$$G_2(s) = \varepsilon(s)/\Theta_C(s)$$

$$G_3(s) = U_P(s)/\varepsilon(s)$$

Figure 4

The significant difference is that $r(j\omega)$ was, in effect, discarded in finding Y_p, but $g_x^T \varepsilon$ is not uncorrelated with x or u_p and must be accounted for in the identification problem.

A Numerical Example

To evaluate the numerical properties and the sensitivity to the a priori selected parameters (V_y, W, V_u, τ) a fast time simulation of the pilot model equations has been assembled, and the simulated control task is shown in Figure 5 . As shown, the task is that of pursuit tracking with $11.7/s^2$ controlled element dynamics, and the displayed command signal is filtered white noise with the filter transfer function given ($\theta_c(s)/w(s)$). The state vector is shown, the known gain vector to be identified is listed, and the weights in the objective function used are given. A sample time history of the state and simulated pilot's control input is depicted in Figure 6. Such time histories were sampled at 10 msec intervals and the gains estimated from time windows of data 5, 10, 15, 20, 25, 30 and 35 seconds wide. The root-sum-squred percent error of the five estimated gains is shown in Figure 7. Where

$$E_{RSS} = \left[\sum_{i=1}^{5} \left[\frac{g_i - \hat{g}_i}{g_i} \right]^2 \right]^{1/2}$$

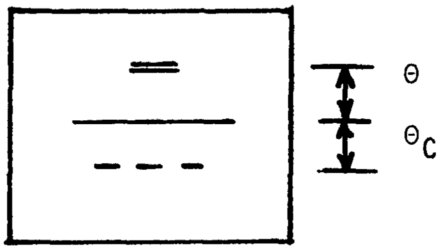
As shown, about 30 seconds of data is required to obtain less than 10% rss error in this example. Other dynamics of higher order, and therefore more gains, will be evaluated in the near future and the convergence will not be as rapid.

The importance of using the proper corrections (e.g., M_{cor} and N_{ν_u}) is shown in Figure 8 , in which the five exact gains, $g_1 \rightarrow g_5$ are shown, along with two sets of gain estimates. The set labeled "uncorrected" was obtained via straight-forward least squares (i.e., M_{cor} and N_{ν_u} not included).

EXAMPLE

PLANT: $\frac{\Theta(s)}{\delta(s)} = \frac{11.7}{s}$

$$\frac{\Theta_C(s)}{W(s)} = \frac{3.67}{s^2 + 3s + 2.25}$$



$\varepsilon = \Theta_C - \Theta$

$$X^T = \begin{bmatrix} \Theta_C & \dot{\Theta}_C & \Theta & \dot{\Theta} \end{bmatrix}$$

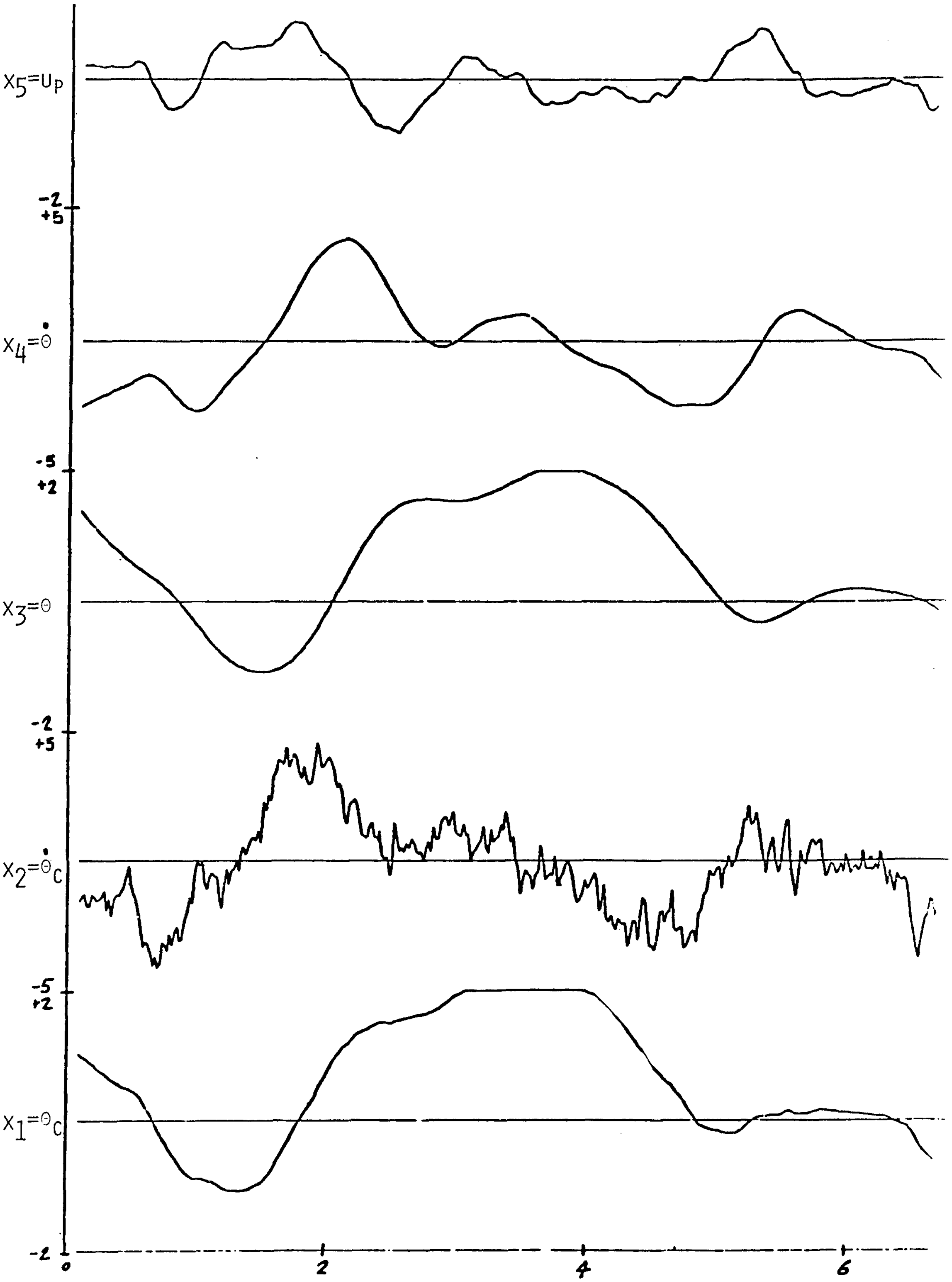
$$\dot{X} = AX + BU + N$$

$$\dot{U}_P = G_X^T \hat{X}_P + G_U U_P + V_U$$

$$\begin{bmatrix} G_X^T & G_U \end{bmatrix} = \begin{bmatrix} 5.53, & 1.86, & -6.76, & -3.69, & -9.28 \end{bmatrix}$$

$Q_\varepsilon = 16/.35$, $Q_{\dot{\varepsilon}} = 1/.35$, $Q_{\dot{\delta}} = 1$

Figure 5



TIME(SEC)

Figure 6 Time Histories

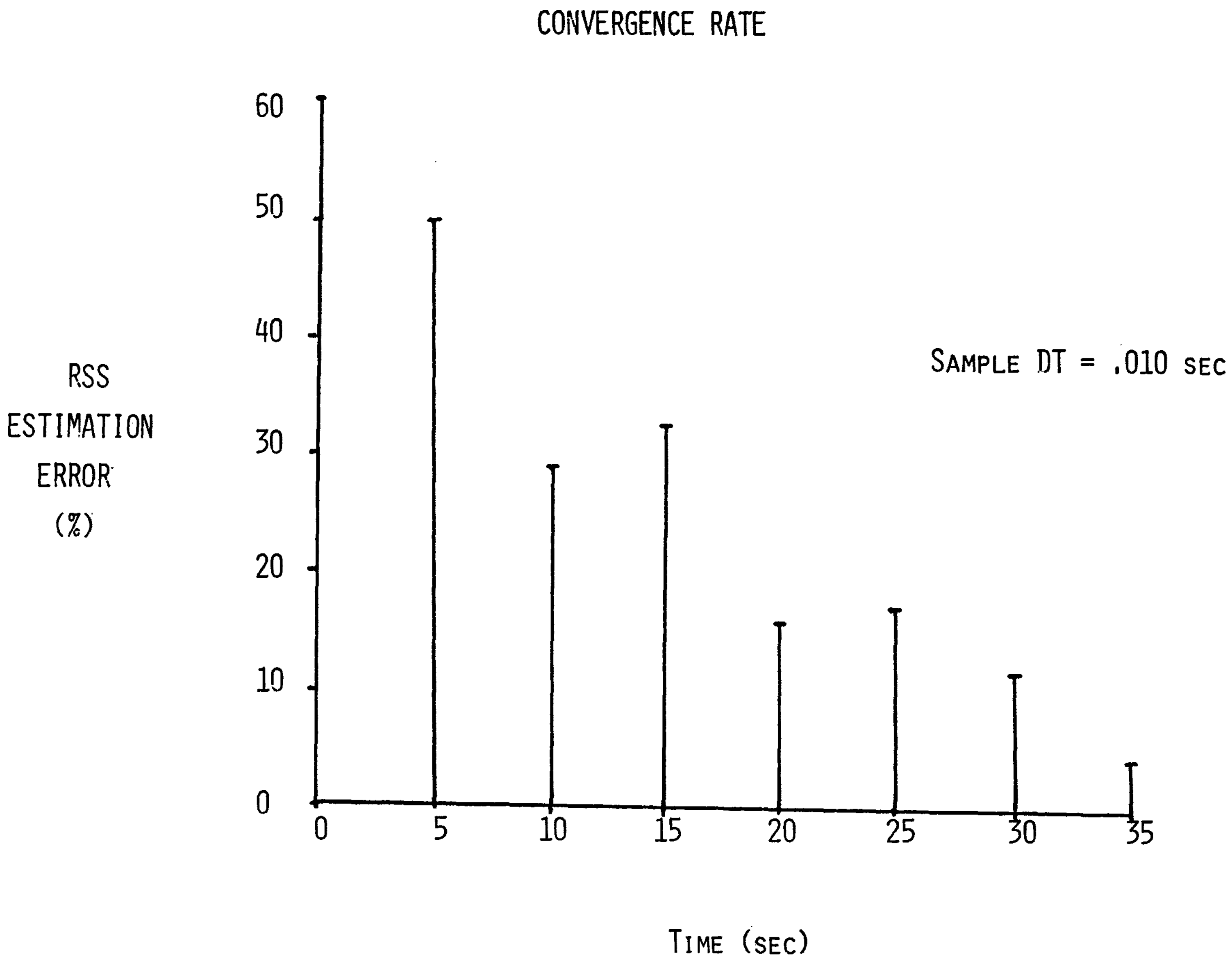
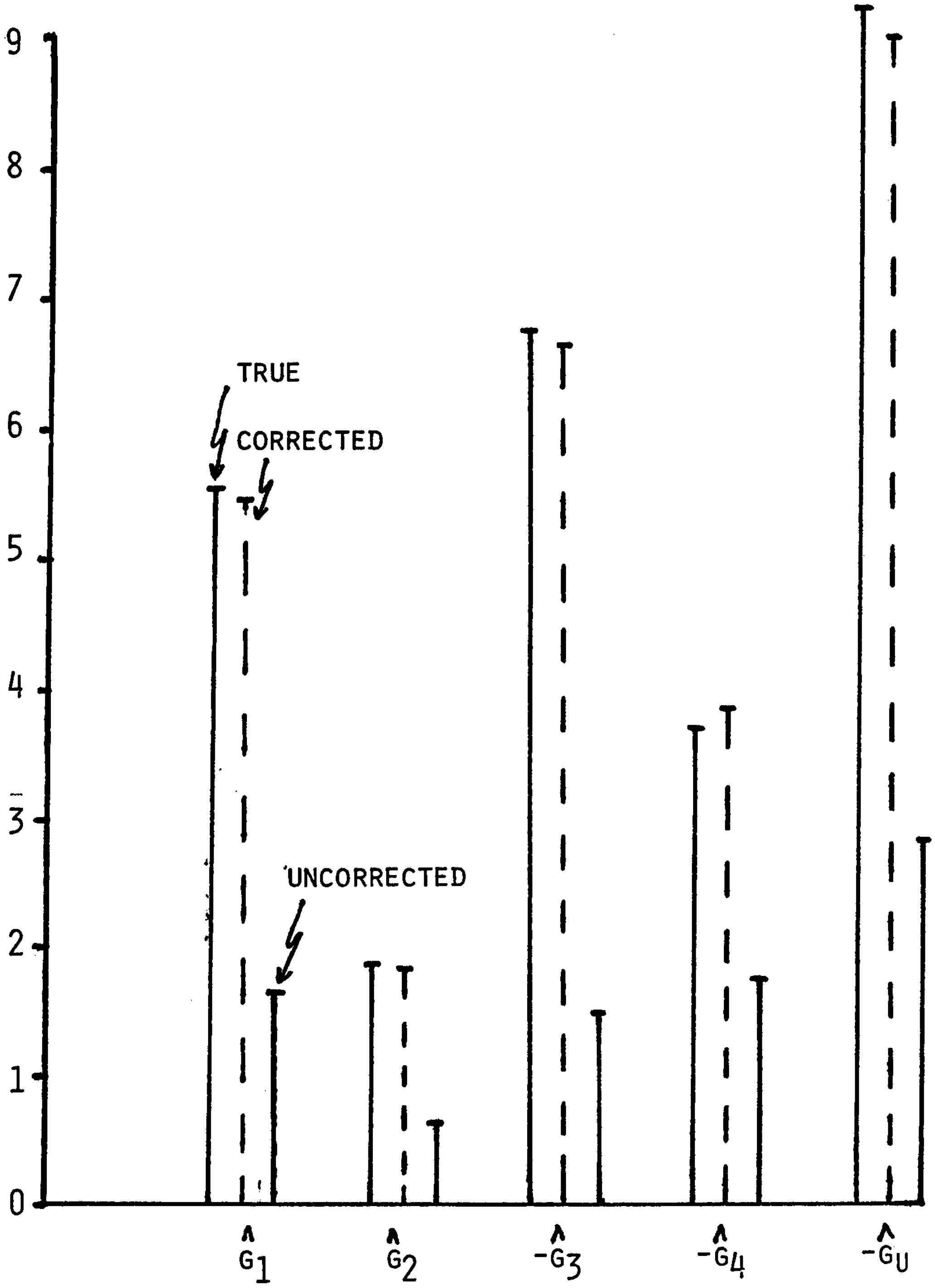

CONVERGENCE RATE
RSS
ESTIMATION
ERROR
(%)
60
50
40
30
20
10
0
0
5
10
15
20
25
30
35
SAMPLE DT = .010 SEC
TIME (SEC)

Figure 7

IMPORTANCE OF CORRECTIONS



GAIN ESTIMATES

Figure 8

Conversely, the "corrected" set used perfectly corrected data, or the actual $\hat{X}$'s in the identification. Both sets of gain estimates are based of 50 seconds of data. Clearly, in this case again, the corrections are important. Further verification of the method is in process.

Inference of the Objective Function

Attention is now turned to estimation of the objective function weightings from the gain estimates just discussed. (Note, this is referred to in the control literature as the "inverse problem".) These weights are related to the gains via the Riccati matrix K, the solution of

$$\tilde{A}^T K + K\tilde{A} + \tilde{Q} - K\tilde{B}G^{-1}\tilde{B}^T K = 0$$

and

$$\left[\begin{array}{c:c} g_x^T & g_u^T \end{array}\right] = -G^{-1}\tilde{B}^T K$$

where

$$\tilde{A} = \left[\begin{array}{c:c} A & B \\ \hdashline 0 & 0 \end{array}\right] \qquad \tilde{Q} = \left[\begin{array}{c:c} C^T Q_y C & 0 \\ \hdashline 0 & R \end{array}\right]$$

$$\tilde{B}^T = \left[\begin{array}{c:c} 0 & I_u \end{array}\right]$$ I_u = identity of dimension equal to u_p control vector

And recall that Q_y, R, and G are the weightings defined in Eqn (0). Now due to the structure of the OCM, we are able to reduce the above into some simpler relations. First, noting that letting $G = I_u$, without loss of generality (at least in the case of scalar control input u_p), we obtain

$$K = \left[\begin{array}{c:c} L & -g_x \\ \hdashline -g_x^T & -g_u \end{array}\right], \quad g_u = g_u^T$$

and

$$R = g_u g_u^T + g_x^T B + B^T g_x \qquad \text{(III.a)}$$

$$0 = g_x g_u - LB + A^T g_x \qquad \text{(III.b)}$$

$$C^T Q_y C = g_x g_x^T - LA - A^T L \qquad \text{(III.c)}$$

Now L can be eliminated in the last two relations if desired, and have

$$B^T C^T Q_y CB = B^T g_x g_x^T B - H - H^T \qquad \text{(IV)}$$

where

$$H = (g_u g_x^T + g_x^T A)AB$$

By observing Equation III (and IV) we can see that if estimates of g_x and g_u are available, and plant and observation matrices A, B, and C are known, the R weighting can be obtained directly, but Q_y requires special attention. From Egns. III.b and .c we see that if L can be obtained by solving III.b, an n x n matrix equation with only C^TQC unknown results from III.c. But this is only possible if B^{-1} exists, which is only true if the number of independent control inputs (in u_p) equals the number of states (in x)-an unlikely situation.

An alternate attack using Eqn. IV leads to similar results. One could conceivably solve for a diagonal Q_y via a numerical method like Newton-Raphson, but that requires the matrix CBB^TC^T to be invertable. This is possible if the number of control inputs (in u_p) equals the number of outputs (or y), (or the system transfer function matrix is square). Although this is less restrictive than the previous situation, it is also untrue in many applications of interest to us here. So the following conclusions may be stated, that in general a unique set of objective function weights may not be obtainable from gain estimates alone. This result is not new, we've just looked at it in the context of our specific problem.

Although improved methods are currently under investigation in this regard, we may always test <u>assumed</u> objective function weights to determine if they're feasible. This is considered a reasonable alternative since in an actual experiment, the analyst knows several reasonable statements for the objective function, and he may at least test them to see which one is best supported by the data. To pursue this approach, the accuracy of the gain estimate will also be developed such that statistical hypothesis tests may be performed. But for now, this is an important consideration.

In the case of our numerical example, Equation III.a leads to R = 0, and Equation IV yields

$$(11.7)^2(q_{\dot{\varepsilon}} + q_{\dot{\theta}}) = -2(11.7)g_u g_{x_3} + (11.7\ g_{x_r})^2$$

where

$$Q_y = \begin{bmatrix} q_\varepsilon & 0 & 0 & 0 \\ 0 & q_{\dot{\varepsilon}} & 0 & 0 \\ 0 & 0 & q_\theta & 0 \\ 0 & 0 & 0 & q_{\dot{\theta}} \end{bmatrix}$$

Using the estimated gains we obtain

$$(q_{\dot{\varepsilon}} + q_{\dot{\theta}}) \cong 2.89 \qquad \text{(actually } q_{\dot{\varepsilon}} \text{ was } 1.35 \text{ and } q_{\dot{\theta}} \text{ was } 0)$$

Now "guessing" that q_θ and $q_{\dot{\theta}}$ were zero, at first, we may iterate on q_ε and solve III.c, then check with III.b. If no $q_\varepsilon > 0$ led to a solution, then the assumption of q_θ and $q_{\dot{\theta}}$ equal to zero would need revision. Finally, note that from Equation III.b, we can actually solve for as many columns (and rows) of L as the number of control inputs (or rank B), and this part of the L matrix may be used to check results from III.c.

Acknowledgement

This work is being performed with the cooperation of NASA Dryden Research Facility under NASA Grant NAG4-1. Mr. Donald T. Berry is the technical monitor, and his support, and that of NASA, is appreciated.

References

1. Kleinman, D., Baron, S., and Levison, W.H., "An Optimal Control Model of Human Response, Part I: Theory and Validation," Automatica, Vol. 6, 1970, pp. 357-369.

2. Hess, R.A., "Prediction of Pilot Opinion Rating Using an Optimal Pilot Model," Human Factors, Vol. 19, Oct. 1977, pp. 459-475.

3. Schmidt, D.K., "On the Use of the OCM's Objective Function as a Pilot Rating Metric," 17th Annual Conf. on Manual Control, UCLA, June, 1981.

4. Levison, W.H., "Methods for Identifying Pilot Dynamics," Proceedings of the USAF/NASA Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, Edwards AFB, Jan. 19-21, 1982.

5. Bryson, A.E., and Ho, Y.C., Applied Optimal Control, Halsted Press, NY, 1975.

6. McGruer, D.T., and Krendel, E.S., Mathematical Models of Human Pilot Behavior, AGARDograph, No. 188, Jan. 1974.

AIAA WORKSHOP - Flight Testing to Identify Pilot Workload and Pilot Dynamics, January 21, 1982, Air Force Flight Test Center, Edwards Air Force Base, CA.

MODELING, DESIGN AND VALIDATION OF AN ADVANCED COCKPIT DISPLAY SYSTEM*

By

Dr. Jonathan Korn
Dr. Sol W. Gully
ALPHATECH, Inc.
3 New England Executive Park
Burlington, MA 01803
And
Dr. David L. Kleinman
University of Connecticut
Storrs, CT

ABSTRACT

Validation of a model-based methodology for the design and evaluation of advanced cockpit information-display systems is undertaken. The technique is applied to determine information and display requirements in an A10 terrain-following flying task. Four candidate display systems, including a flight director system, are proposed and rank-ordered across dimensions of workload and performance. Validation of the analytical predictions is accomplished through man-in-the-loop simulation experiments. It is concluded that the methodology, which can determine the limit information to the best quantities needed by the pilot to perform effectively, can be a valuable tool in the development of advanced cockpit information systems.

INTRODUCTION

The increased complexity of advanced tactical air-to-air/close air support aircraft has elevated the pilot to a role of systems manager. The pilot must allocate sensors, evaluate threats, select weapons, employ ECM, etc., in addition to the omnipresent task of flying the aircraft. The increase in the number of systems and subsystems has required the display of increasingly more information to the pilot about his aircraft and the environment. Unfortunately, this has increased the pilot's workload. A methodology, therefore, is required to determine and limit information to the best informational set needed by the pilot to perform his various tasks effectively.

If pilot and aircraft are to function as an effective combat system, it is essential that the workload associated with the basic flying task be reduced through either display or control automation. Thus, the pilot would be in a

---

* This work performed under the following contract to AMRL: F33615-80-C-0528

position where he can allocate a greater fraction of his capacity for decision making relative to weapon assignment, etc. Nowhere is this problem more critical than in high-speed, low-level, adverse-weather interdiction where the workload demands of the flying tasks are severe.

The design of automated control/display systems for reducing workload in piloting tasks is generally accomplished through extensive recourse to man-in-the loop simulations. This can be an expensive and time-consuming approach, especially when large numbers of competing designs and/or parameter sets must be evaluated and iterated. While simulation is ultimately necessary in the development of control/display systems, it would be desirable to have a computer-based tool that could perform a preliminary evaluation on a wide variety of systems on a relative performance basis. In this manner, one would only need to retain those systems exhibiting promise for further evaluation by manned simulation.

A model/computer-based methodology has recently been proposed [1-2] that offers considerable potential for application to tactical aircraft display systems design and evaluation. Although this methodology has been applied to rank-order control/display systems for a CH-47[1], the design technique has remained largely unvalidated due to the lack of subsequent manned simulation to assess the preliminary analytic results. In this study the design procedure, as proposed by Kleinman and Curry, is extended to a practical cockpit scenario. Subsequently, we validate the model predictions via man-in-the-loop simulation experiments. The piloting control task considered is a representation of a zero-visibility, high-speed, low-level terrain following scenario with an A10 aircraft. The aircraft dynamics are approximated by their short-period longitudinal equations in the analysis and simulation.

The focus of this study is on display, design, and evaluation, including optimal synthesis or aggregation of information states, for the pilot control task. The design methodology, when applied to display systems evaluation, follows a three-level procedure. At the first, or information level, the methodology determines the relative importance of each system variable to closed-loop task performance and determines the optimal synthesis of information in the contex of a flight director signal. At the second, or display element level, the information requirements are integrated to propose several different realistic display systems. Human generated information processing limitations are included at this level, and for each candidate design, performance versus workload curves are generated. The different display systems are then rank-ordered across dimensions of performance and workload. The proposed displays that resulted via the application of the methodology to the terrain following scenario included: (1) a terrain predictor/flight path vector display, (2) a tunnel display, (3) an integrated tunnel-predictor display, and (4) an optimally designed flight director display.

The display format level is the third and final phase of the design process. Here, specific display formats are suggested for the presentation of the candidate display systems, guided by the sensitivities, attentional demands, etc., that are predicted at the display element level. The format level is the precursor to manned simulation experiments. In this effort, a man-in-the-loop experimental program was conducted at the University of Connecticut to evaluate the four candidate display systems. The objective of these experiments was to validate the overall display design procedure, including the analytic rank-ordering, the workload assessment, and the flight director design synthesis processes.

The analytical nucleus to the display design methodology is the Optimal Control pilot Model (OCM)[3]-[5]. A detailed description of the OCM, in conjunction with the three-level display design process, is given in [6]. In this paper we first introduce the control task of interest, including the A10 aircraft dynamics and the terrian profile characteristics. We apply then the display design procedure, which yields the major analytical results of this study, including the predictions of attentional allocation, workload demands, and performance rank-ordering. Finally, the experimental program is described, and the experimental results discussed.

APPLICATION OF THE CONTROL TASK TO THE OCM

A-10 LONGITUDINAL AIRCRAFT DYNAMICS

As indicated, the control task of interest is high-speed terrain following for a representative low-level attack flight condition. The basic set of longitudinal equations being used is the short-period dynamics. These are the perturbation equations written about straight and level flight, and they describe the aircraft longitudinal rotations about its center of mass.

The (two-degree of freedom) transfer functions of interest are [7]

$$\frac{\alpha}{\delta}(s) = \frac{1}{U_o}\,\frac{Z_\delta s + (U_o M_\delta - Z_\delta M_q)}{\Delta(s)} \tag{1}$$

$$\frac{q}{\delta}(s) = \frac{(M_\delta + Z_\delta M_{\dot{w}})s + (Z_\delta M_w - M_\delta Z_w)}{\Delta(s)} \tag{2}$$

$$\Delta(s) = s^2 - (U_o M_{\dot{w}} + Z_w + M_q)s + (M_q Z_w - U_o M_w) \tag{3}$$

where q = aircraft pitch rate, α = angle of attack, U_o = nominal air speed, δ = elevator angle, and Z_δ, M_δ, M_q, $M_{\dot{w}}$, Z_w are the pertinent stability derivatives.

In the present work the stability derivatives were derived from data currently used on the Aerospace Systems Division A-10 nonlinear hybrid simulation, and, the numerical values are given in [6]. The nominal air speed U_o = 468 ft/sec.

In addition, the aircraft pitch angle, θ, and its altitude, h, are necessary in the modeling process. Therefore

$$\dot{\theta} = q \quad ; \quad \dot{h} = U_o(\theta - q) \tag{4}$$

TERRAIN PATH MODEL

The next step is to select an appropriate terrain path model that serves as the excitation "signal" to the system. An easily implemented one dimensional model for the terrain height Π as a function of time consists of a Markov process passed through a 2nd order filter*[6]. The Markov process is generated by

$$\dot{z}(t) + \frac{2U_o}{\lambda} z(t) = \frac{2U_o}{\lambda} \xi(t) \tag{5}$$

where $\xi(t)$ is a white Gaussian noise, and λ represents the spatial terrain variations. Next, $z(t)$ is filtered through

$$H(s) = \frac{\omega^2 U_o^2}{s^2 + 2\zeta\omega U_o s + \omega^2 U_o^2} \tag{6}$$

where $\omega = 2\pi/D$ is the natural frequency of the terrain (D = "terrain period") and ζ is the damping ratio. In order to represent a realistic terrain profile, the terrain parameters are selected appropriately [6]. The parameter values chosen to give the most appealing terrain characteristics where

$$D = 10000' \quad ; \quad \lambda = D/\pi \quad ; \quad \zeta = 1/\sqrt{2}$$

The white noise ($\xi(t)$) intensity was chosen to yield a reasonable terrain RMS value of ~136 ft.

OCM APPLICATION

The aircraft dynamics and the terrain states are combined into a single state equation as required by the OCM. The augmented state is defined as

$$x' = [\Pi \ \dot{\Pi} \ \ddot{\Pi} \ \alpha \ q \ \theta \ e] \tag{7}$$

where $e(t) \overset{\Delta}{=}$ flight path error = $\Pi(t) - h(t)$. The total state space equation is now

$$\dot{x}(t) = Ax(t) + B\delta(t) + E\xi(t) \tag{8}$$

---

*The term Π should represent a spatial rather than temporal terrain model. However, assuming a constant velocity, U_o, and neglecting the flight path flactuations, we can write $\Pi(r) \simeq \Pi(U_o t) = U_o \Pi(t)$, where $r \simeq U_o t$ represents the traveled distance.

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{-2U_o{}^3\omega^2}{\lambda} & -\left(\frac{4\zeta\omega}{\lambda} + \omega^2\right)U_o{}^2 & -\left(2\zeta\omega + \frac{2}{\lambda}\right)U_o & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & Z_w & 1 & 0 & 0 \\ 0 & 0 & 0 & U_o(Z_w M_{\dot{w}} + M_w) & U_o(M_{\dot{w}} + M_q) & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & \frac{\pi}{180°} U_p & 0 & \frac{-\pi}{180°} U_o & 0 \end{bmatrix}$$

$$B' = \begin{bmatrix} 0 & 0 & 0 & \frac{Z_\delta}{U_o} & M_\delta + Z_\delta M_{\dot{w}} & 0 & 0 \end{bmatrix}$$

$$E' = \begin{bmatrix} 0 & 0 & \frac{2U_o^3\omega^2}{\lambda} & 0 & 0 & 0 & 0 \end{bmatrix}$$

The numerical values of A, B, E are:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -.025 & -.21 & -.71 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1.37 & 1 & 0 & 0 \\ 0 & 0 & 0 & -3.08 & -1.68 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 8.17 & 0 & -8.17 & 0 \end{bmatrix}$$

$$B' = \begin{bmatrix} 0 & 0 & 0 & -0.06 & -6.84 & 0 & 0 \end{bmatrix}$$

$$E' = \begin{bmatrix} 0 & 0 & .025 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The terrain following task requires the pilot to follow the path-over-terrain $\Pi(t)$, as closely as he can, subject to his inherent limitations. Thus, the quantity he should minimize is the flight path error (FPE) $e(t)$. In addition, the pilot should avoid large vertical acceleration values, or g's. These requirements on $e(t)$ and $g(t)$ (or, equivalently on $q(t)$) are expressed in the cost functional associated with the OCM, with the pilot's subjective weightings q_e and q_q, viz.,

$$J(\delta,f) = E\{q_e e^2(t) + q_q q^2(t) + q_{\dot{\delta}} \dot{\delta}^2(t)\} \quad , \tag{9}$$

where the control rate weighting, $q_{\dot{\delta}}$, is determined by the pilot's neuromuscular time constant τ_N [3]-[6]. The cost functional $J(\delta,f)$ is minimized with respect to the control δ and the attentional allocation vector f.

The nominal altitude of the trajectory above the terrain ($\Pi(t)$) the aircraft must follow is 200 ft., and we assume that the maximal excursion, or FPE, should be no more than 20%, or $e_{max} = 40$ ft. Also, we assume that the maximal vertical acceleration to be tolerated by the pilot is $a_{z,max} = 1g = 32.2$ ft/sec^2. Thus, we compute the FPE and pitch rate weighting according to [1],[2]-[6], viz.,

$$q_{max} \cong \frac{180°}{\pi} \cdot \frac{a_{z,max}}{U_o} \cong 4°/\text{sec}$$

$$q_e = \frac{1}{e_{max}{}^2} = \frac{1}{40^2} = 6.25 \cdot 10^{-4}\text{ft}^{-2}$$

$$q_q = \frac{1}{q_{max}{}^2} = \frac{1}{4^2} = 6.25 \cdot 10^{-2}\ \text{deg}^{-2}\ \text{sec}^2$$

The assignment of the neuromuscular time constant $\tau_N = .15$ sec completes the specification of the cost functional Eq. (9). It must be indicated that J is the "nominal" cost functional. With new display systems incorporated in the information set, it is possible that the pilot will attempt to minimize additional indicators. This will be explained in detail in the sequel.

DISPLAY SYSTEMS DESIGN

We present now the major analytical results of the display systems design for the candidate terrain-following task.

INFORMATION LEVEL ANALYSIS

The information level analysis is used to determine information requirements in the preliminary stages of display design and is the cornerstone to the subsequent display level analysis, There are two basic tasks at the information level: the information requirements assessment and the flight-director design.

Information Requirements

At this step we assume that all state variables, and possibly their rates are accessible to the pilot. Only the plausible rate variables should be considered. After some preliminary analysis, the following set of variables was explicity considered:

$$y' = [\Pi\ \dot{\Pi}\ \ddot{\Pi}\ e\ \dot{e}\ h\ \dot{h}\ \theta\ q] \qquad (10)$$

In the OCM application we use the cost functional defined by Eq. (9). The resulting terrain following performance, in terms of absolute values of the flight path error, is not of immediate interest at this level. We are interested in the relative importance of each state variable to the control task. Thus, the performance index (9) is minimized subject to $\Sigma f_{ci} = f_c$, $f_{ci} \geq 0$, where f_{ci} is the attentional allocation to the ith indicator. This procedure is repeated for several values of control attention, f_c, and the summation is taken over the nine elements of Eq. (10). The optimal f_{ci} that ensue provide the relative importance of each of the states, i.e., the information requirements for the control task.

The OCM analysis yields the relative attentional allocation (f_{ci}/f_c), summarized in Table 1 and Figure 1. The attentional allocations are examined at four levels of total control attention, $f_c = 0.9, 0.7, 0.5, 0.3$. The only variables that command significant pilot attention are the terrain "vertical acceleration" $\ddot{\Pi}(t)$, the FPE rate, $\dot{e}(t)$, and the FPE $e(t)$. All other system states demand negligible attention in the pilot model.

TABLE 1. RELATIVE ATTENTIONAL ALLOCATION-INFORMATION LEVEL

f_c	$f_{\ddot{\Pi}}/f_c$	$f_{\dot{e}}/f_c$	f_e/f_c
0.9	0.46	0.34	0.1
0.7	0.46	0.36	0.1
0.5	0.42	0.36	0.1
0.3	0.37	0.33	0.1

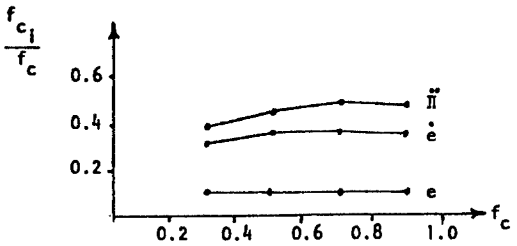


Figure 1. Relative Attentional Allocation-Information Level.

Notice that there is a consistency in these results, i.e.,

$$\frac{f_e}{f_c} = \text{const.}, \quad \frac{f_{\ddot{\Pi}}}{f_c} \cong \text{const.}, \quad \frac{f_{\dot{e}}}{f_c} \cong \text{const.}$$

Thus, the relative importance of the key state variables does not change over different levels of control attention f_c.

The results, which indicate that $\ddot{\Pi}(t)$, $\dot{e}(t)$, and $e(t)$ is the critical information base, must now be interpreted. A major design issue, at this point, is in determining whether or not a separate display of each critical variable is required for use by the pilot. It is clear that a FPE indicator should be included since it is essential that the pilot knows his instantaneous position with respect to the desired path $\Pi(t)$. Such a display is easily constructed using, e.g., a radar altimeter. Also, the pilot can easily extract the FPE rate from a well designed display of $e(t)$, and therefore, a separate $\dot{e}(t)$ indicator is not needed.

The major portion of the pilot's attention is allocated to $\ddot{\Pi}(t)$. In practice, this variable is difficult to derive and display explicity. Also, it is not clear how such a display might be used by the pilot. One must, therefore, synthesize a display in which $\ddot{\Pi}(t)$ will be incorporated in a practical manner.

One interpretation of the requirement for rate information $\ddot{\Pi}$(and $\dot{e}$) is the necessity to display the path-over-terrain at a future time, i.e., at some distance ahead. Such displays will be synthesized, modeled, and compared at the display element level.

Flight Director Design

The basic concept of a flight director is to provide the pilot with (synthesized) information that is useful for control, thus rendering the piloting task easier in some sense. This section describes the flight director design procedure that uses the quadratic synthesis technique of the OCM to determine the optimal (linear) aggregation of system states to be used as a display.

The flight director signal is a linear combination of the system states,

$$y_{FD}(t)=h'x(t) \qquad (11)$$

The gains h are chosen so that if $y_{FD}(t)$ is kept "small" by the pilot, the resulting aircraft motion will be desirable. Since the pilot is in control of the vehicle, there are two issues that relate to the harmony between $y_{FD}(t)$ and pilot response. The first concerns the nature of the control task as viewed by the pilot. Thus, the task of keeping $y_{FD}(t)$ small should not conflict with the overall pilot-control task requirements. The second issue relates to the required form of the pilot compensation, as y_{FD} and δ are in one-to-one correspondence. From a reduced workload point of view, one should design a flight director signal $y_{FD}(t)$ such that the transfer function from input $\delta(t)$ to $y_{FD}(t)$ is approximately K/s. The required pilot compensation would then be simple proportional feedback

$$\delta(t)\approx K \cdot y_{FD}(t) \qquad (12)$$

From the OCM, the pilot's control strategy is given by

$$\tau_N\dot{\delta} + \delta = -L\hat{x}(t) + v_\delta(t) \qquad (13)$$

where $\hat{x}(t)$ is the state estimate, $v_\delta(t)$ is a white motor noise, and the gains L (and τ_N) are obtained by minimizing the cost function $J(\delta)$ that is associated with the terrain following task, Eq. (9). To begin with, we suggest the obvious design choice

$$y_{FD}(t) = Lx(t) \quad , \qquad (14)$$

i.e., h' = L. Such a design, however, does not consider the possibility that the flight director, once added to the display panel, modifies the pilot's control task and hence changes the cost functional $J(\delta)$. Excluding y_{FD} from the cost functional implies that the pilot's control objectives are basically the same as before introducing this signal. This is not a reasonable assumption. Indeed, including the y_{FD} with $J(\delta)$, in addition to the other terms, implies that one of the pilot's direct control objectives is to keep the y_{FD} small. Thus, we assume that the director signal y_{FD} is explicitly controlled.

The control cost functional, modified to weight deviations of $y_{FD}(t)$ is now

$$J'(\delta) = J(\delta) + E\left\{q_{FD}y_{FD}^{2}(t)\right\} \tag{15}$$

The weighing term q_{FD} is selected

$$q_{FD} = \frac{1}{y_{FD,max}^{2}} \tag{16}$$

to be consistent with the choice of the q_{yi}'s. The maximal flight director excursion is computed according to the rule

$$y_{FD,max} = \sum_{i} \gamma_i |l_i| x_{i,max} \tag{17}$$

where the l_i are the entries of the gain vector L, and the γ_i is 0 or 1 to indicate which variables are of concern in forming $y_{FD,max}$. We select

$$\gamma_i = \begin{cases} 1 \text{ if } x_i \text{ is a positional variable} \\ 0 \text{ if } x_i \text{ is a rate variable} \end{cases} \tag{18}$$

Thus, the flight director signal is at its maximum value when all error displacements are at their design limits.

With the pilot cost functional modified as in Eq. (15), the pilot model control is now obtained by minimizing a new cost functional, the result being

$$\tau_N \dot{\delta} + \delta = -L\hat{x}(t) + v_{\delta}(t) \tag{19}$$

But since the cost functionals of Eq. (9) and (15) are not the same, the optimal gains L in Eq. (19) differ from those in Eq. (13). Hence, the flight director signlas of Eq. (14) and the required pilot control gains in Eq. (19) are no longer in harmony. This mismatch can be corrected via the iterative process of computing feedback gains and flight director signals as shown in [1].

This algorithm has given rapid convergence in the terrain following problem. Only four iterations have been needed, and the resulting converged gains were within 10 percent of the initial gain values obtained from Eq. (9).

The numerical results of the flight director design process are:

1. design parameter, $y_{FD,max} \simeq 3$
2. flight director weighting, $q_{FD} = \frac{1}{y^2_{FD,max}} = \frac{1}{3^2} \simeq .1$
3. flight director signal

$$y_{FD} = Lx = \left[3 \cdot 10^{-3} \ .13 \ .11 \ .76 \ -.48 \ -1.2 \ .06 \right] x$$

We do not yet examine terrain following performance with the flight director display, as this is the subject of the display element level in wich all proposed displays are compared.

DISPLAY ELEMENT LEVEL ANALYSIS

The next step in the display design methodology is to evaluate control performance for several candidate display systems. The evaluation process begins with the definition of a performance metric for the terrain following task.

Often, the control performance requirements of the pilot-vehicle combination are specified in terms of allowable excursions or desired RMS deviations in system states. The design specifications are generally a function of mission requirements or flight conditions. Recall that the nominal altitude of the aircraft over the terrain is 200', and that the maximal FPE deviation "allowed" is 40'. Thus, $e_{max} = 40'$ is chosen as the design tolerance in the control performance metric, which is defined as

$$P_c = \int_{-e_{max}}^{e_{max}} \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp \left\{ \frac{e^2(t)}{2\sigma_e^2} \right\} d[e(t)] \tag{20}$$

This performance metric measures the probability that the aircraft does not deviate more than $e_{max} = \underline{+}40'$ from the desired path $\Pi(t)$. The maximum level of control performance is selected as $P_{c,max} = .99$.

The OCM parameters to be used in all candidate display systems are:

observation noise (for all eventual indicators), $\rho_{yi} = .20$ dB;
motor noise, $\rho_\delta = -25$ dB; time-delay, $\tau_D = .15$ sec.

The neuromuscular time-constant has already been specified as $\tau_N = .15$ sec.

Using these parameter values and the performance metric P_c, Eq. (20), we now evaluate and compare the candidate display systems.

Display System Synthesis

STATUS DISPLAY

The status display system consists of a FPE $e(t)/\dot{e}(t)$, and pitch and pitch rate, $\theta(t)$, $q(t)$, indicators. This rudimentary display is not a truely synthesized system at which this study is aimed; nevertheless, we use the status display as a benchmark in the display system comparison process. The selection of $[e, \dot{e}, \theta, q]$ for the status display set is natural [6], and all subsequent display systems include the status display set. Naturally, the status display system alone should yield the worst terrain following performance. Also, we assign indifference threshold values, a_i to $e(t)$, $\dot{e}(t)$, $\theta(t)$, and $q(t)$, according to the rule [6] $a_i = |y_{i,max}|/8$. Following the usual assumption that rate variable thresholds are one half of the thresholds on the corresponding position variables, we obtain

$$a_e = \frac{1}{8} e_{max} = \frac{1}{8} \cdot 40' = 5 \ ; \quad a_{\dot{e}} = \frac{1}{2} a_e = 2.5'/sec \ ;$$

$$a_q = \frac{1}{8} q_{max} = \frac{1}{8} \cdot 4°/sec = 0.5°/sec \ ; \quad a_\theta = 2 \cdot a_q = 1°$$

PREDICTOR DISPLAY

Grunwald and Merhav have shown [8]-[9] that acceleration cues are vital in visual field control and that they are obtained by estimating the future vehicle path. These findings are in close agreement with our information level analysis, in which it was shown that the terrain "acceleration", $\ddot{\Pi}(t)$, demands the largest amount of attention among all system states. Such information can best be derived (or estimated) by the pilot when the vehicle's future altitude is displayed relative to the future terrain-path. We define T as the prediction time (i.e. $T \simeq D_o/U_o$, where D_o is the distance ahead at which the prediction is displayed), and the predictor signal is then given by

$$e_p(t) = \Pi(t+T) - h(t+T) \qquad (21)$$

The underlying assumption in the predictor display is that the predictor vehicle path is along the velocity vector $\vec{U}_o$. We rewrite Eq. (21) to reflect this assumption, viz.,

$$e_p(t) = \Pi(t+T) - \left[h(t) + T\dot{h}(t)\right] = \Pi(t+T) - h(t) - \frac{\pi T U_o}{180°}\left[\theta(t) - \alpha(t)\right] \qquad (22)$$

Also, the OCM assumes that the pilot derives rate information from the predictor indicator, $\dot{e}_p(t)$. The appropriate equation for this signal is

$$\begin{aligned} \dot{e}_p(t) &= \dot{\Pi}(t+T) - \dot{h}(t) - T\ddot{h}(t) \\ &= \dot{\Pi}(t+T) - \frac{\pi U_o}{180°}\theta(t) + \frac{\pi U_o}{180°}\left[1+TZ_w\right]\alpha(t) + \frac{\pi T Z_\delta}{180°}\delta(t) \qquad (23) \end{aligned}$$

Such a display is easy to simulate and implement using e.g., a forwardlooking radar. An actual display that was used in the experimental validation is shown in Figure 6. It is important to indicate that the projection of $\vec{U}_o$ on a normal surface located T seconds flight time ahead is required (represented by the cross in Figure 6.

Equations (22-(23) in their present form cannot be modeled directly in the OCM steady-state analysis, where only events at time t are treated. This problem can be "solved" by replacing the signals $\Pi(t+T)$ and $\dot{\Pi}(t+T)$ with their (optimally) predicted future values, viz.,

$$\hat{\Pi}(t+T) = E\left\{\Pi(t+T) \mid \Pi(t),\ \dot{\Pi}(t),\ \ddot{\Pi}(t)\right\} \ ; \quad \hat{\dot{\Pi}}(t+T) = \left\{E\ \dot{\Pi}(t+T) \mid \Pi(t), \dot{\Pi}(t), \ddot{\Pi}(t)\right\} \tag{24}$$

Such an approximation can be easily obtained from the 3 x 3 upper-left block of A in Eq. (8). This results in estimates which are a linear combination of the terrain states $\Pi(t)$, $\dot{\Pi}(t)$, $\ddot{\Pi}(t)$ [6]. Specifically we may write

$$\begin{aligned} \hat{\Pi}(t+T) &= p_{\Pi}(T)\Pi(t) + p_{\dot{\Pi}}(T)\dot{\Pi}(t) + p_{\ddot{\Pi}}(T)\ddot{\Pi}(t) \\ \hat{\dot{\Pi}}(t+T) &= r_{\Pi}(T)\Pi(t) + r_{\dot{\Pi}}(T)\dot{\Pi}(t) + r_{\ddot{\Pi}}(T)\ddot{\Pi}(t) \end{aligned} \tag{25}$$

where, as indicated, the p and r coefficients are a function of the prediction time, T. Thus, we rewrite Eqs. (22)-(23) as a linear combination of the system state:

$$\begin{bmatrix} e_p(t) \\ \dot{e}_p(t) \end{bmatrix} = C_p(T)x(t) + D_p(T)\delta(t)$$

$$= \begin{bmatrix} p_{\Pi}(T)-1 & p_{\dot{\Pi}}(T) & p_{\ddot{\Pi}}(T) & \frac{\pi T U_o}{180°} & 0 & \frac{-\pi T U_o}{180°} & 1 \\ r_{\Pi}(T) & r_{\dot{\Pi}}(T) & r_{\ddot{\Pi}}(T) & \frac{\pi U_o}{180°}(1+TZ_w) & 0 & \frac{\pi U_o}{180°} & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ \frac{\pi T Z_\delta}{180°} \end{bmatrix} \delta(t) \tag{26}$$

The indifference thresholds on the predictor are assumed identical to those of the FPE, $e(t)/\dot{e}(t)$.

The last issue which must be addressed is the selection of the prediction time T. The present A10 HUD uses the value $T \simeq 4$ sec. This value has been selected for use in the present display analysis and in the subsequent experiments. It has been shown [6], however, that the optimal prediction time for the given control task and assumed terrain characteristics is $T^* = 1.5$ sec. The numerical values of $C_p(T)$ and $D_p(T)$ of Eq. (26) for $T = 4$ sec. and $T=T^*=1.5$ sec. are given in the Appendix.

THREE DIMENSIONAL PERSPECTIVE TUNNEL DISPLAY

The idea of a displaying 3-D perspecitve tunnel which envelopes the trajectory over the terrain is not entirely new. It has recently been studied and simulated by Grunwald in a helicopter approach context [10]. Such a display is a computer-generated, "through-the-windshield" perspective view of a tunnel that follows the contours of the terrain. In practice, it seems as if the tunnel was flying towards the observer. The pilot, on his part, tries to maintain the aircraft as close to the tunnel's center as possible. Figure 7 shows the tunnel geometry and display which was used in the experimental validation.

The present FPE, e(t), is not explicity available to the pilot from the tunnel display. He can, however, derive sufficient rate and acceleration information, as dictated by the information level results, from a continuance of future flight path errors displayed by the perspective tunnel. It is necessary now, to translate the information provided by the tunnel into an analytical model for application of the OCM.

A plausible approach to the modeling problem has been suggested by Tomizuka [11] in the so-called "finite preview" problem. If the tunnel is sufficiently long, we may assume an infinite preview time. The finite preview problem is then reduced to a common optimal tracking problem, and can be treated as such.

Both approaches, albeit plausible, require major modifications in the OCM methodology and, therefore, are not considered here. The modeling approach taken in the present study treats the tunnel display in the OCM framework. As indicated, the tunnel provides a continuance of future flight path errors, assuming a straight flight path. Formally, this information base may be represented as $e(t+\tau)$, $\tau\varepsilon[o,t_p]$ where, in general, $t_p<\infty$. However, such a representation is impractical from a modeling point of view. One approximation is to replace $e(t+\tau)$ with N_p "indicators" which represent the FPE values at distinct points in the future, viz., $e(t+\sigma_i)$, $\sigma_i = t_p i/N_p$ where i varies from 1 to N_p. The indicators $e(t+\sigma_i)$ would then be treated as independent N_p predictor display systems. Such a model, although requiring an extensive set of equations, if N_p is chosen to be large, can be implemented in the OCM framework. This modeling approach is carried further by assuming that the pilot concentrates on a single distance ahead, i.e., $N_p=1$. Since the underlaying assumption of the OCM is that the pilot adopts an optimal control policy, it is equally valid to assume that, in the tunnel display, he choses the optimal prediction distance/time, T*, when looking down the tunnel path. We, therefore, replace the N_p $e(t+\sigma_i)$ observations with a single indicator $e(t+T^*)$. The tunnel display model, $\kappa(t)$, is then simply

$$\begin{bmatrix}\kappa(t)\\ \dot{\kappa}(t)\end{bmatrix} = \begin{bmatrix}e(t+T^*)\\ \dot{e}(t+T^*)\end{bmatrix} = C_p(T^*)x(t) + D_p(T^*)\delta(t) \qquad (27)$$

where the numerical values of $C_p(T^*)$ and $D_p(T^*)$ $T^*=1.5$ sec. are given in the Appendix.

Despite initial similarities, there is a fundamental difference between a simple optimal predictor display where $T=T^*=1.5$ sec., and a tunnel display. Although both displays are represented by the same equation (26), the tunnel display does not include the velocity vector's tip, $\vec{U}_O$ (Fig. 6) projected on the normal surface located T* seconds flight time ahead. The end point of $\vec{U}_O$ is estimated rather than displayed in the perspective tunnel. This fact is reflected in the OCM by large observation thresholds on $\kappa(t)$, $\dot{\kappa}(t)$. Given the $\pm 40'$ tunnel dimensions, as implemented in the subsequent experiments, we assume

$$a_\kappa = 20', \; a_{\dot{\kappa}} = \frac{1}{2} a_\kappa = 10'/\text{sec}$$

These thresholds are significantly larger than the indifference thresholds used for the simple predictor display (5 ft and 2.5 ft/sec respectively). Naturally, such large thresholds tend to degrade terrain following performance. To overcome this problem the following display system is suggested.

INTEGRATED TUNNEL/VELOCITY VECTOR DISPLAY

In this system we simply superimpose the velocity vector's ($\vec{U}_O$) trace on the existing tunnel display as shown in Fig. 8. Again, the value used is T=4 sec and not T*, in accordance with the current A10 display system. It is obvious that incorporating this new information will reduce the thresholds $a_\kappa/a_{\dot{\kappa}}$, since the pilot now has a reference point about which he will "center" $\Pi(t+T^*)$. Since $\vec{U}_O$ is projected at T=4 sec., and the pilot's "focus" in the tunnel is at T=1.5 sec., the thresholds $a_\kappa/a_{\dot{\kappa}}$ are not reduced to the $a_e/a_{\dot{e}}$ values, as in the predictor display, but rather to an intermediate value. We select

$$a_\kappa = 10', \; a_\kappa = 5'/\text{sec}$$

The information base now includes both tunnel $\kappa(t)/\dot{\kappa}(t)$, and a 4 second predictor, $e_P(t)/\dot{e}_P(t)$, (in addition to the status display). The $e_P, \dot{e}_P$ threshold values remain unchanged, 5', 2.5'/sec. respectively.

FLIGHT DIRECTOR SYSTEM

The status information base is now augmetned with the flight director position and rate observation $y_{FD}(t)$, $\dot{y}_{FD}(t)$ as discussed previsouly. The $y_{FD}/\dot{y}_{FD}$ observation equations are given by

$$y_{FD}(t) = Lx(t) = [3.10^{-3} \; .13 \; .11 \; .76 \; -.48 \; -1.2 \; .06] \; x(t) \qquad (28)$$

$$\begin{aligned} \dot{y}_{FD}(t) &= L\dot{x}(t) = LA_O x(t) + LB_O \delta(t) \\ &= [-3.10^{-3} \; .03 \; .05 \; .91 \; .31 \; -.48 \; 0] x(t) + 3.2\delta(t) \end{aligned} \qquad (29)$$

Also, using the fact that $y_{FD,max} \simeq 3$, the indifference thresholds are

$$a_{FD} = \frac{1}{8} \cdot y_{FD,max} = \frac{1}{8} \cdot 3 \simeq .4; \; a_{FD} = \frac{1}{2} a_{\dot{FD}} \simeq .2$$

Implementation of the flight director in a practical manner, using future terrain path information in lieu of $\dot{\Pi}$ and $\ddot{\Pi}$, is described in the sequel.

Thus, we have proposed and obtained analytical models for five candidate display systems. The next task in the analysis procedure is to evaluate control performance and attention allocation for these systems.

Control Performance: Modeling Results

The performance of each of the display systems is evaluated in terms of

1. control performance, and
2. acceleration stress levels

Following [1]-[6], we introduce now the concept of control and monitoring workload. The control workload metric is based on the fractional attention the pilot allocates among the various display indicators. It is assumed that a pilot distributes a total amount of attention, or workload, $f_T \approx 0.8 < 1.0$ between the tasks of control and monitoring, leaving about 20 percent of his capacity for other duties (e.g., communications). Let f_c and f_m denote, respectively, the control and monitoring attentions, or workloads. Thus. $f_c + f_m = f_T$. The attention allocated for control, f_c, is distributed among all of the display variables $y_1, y_2, \ldots, y_{Ny}$, where y_i and $y_{i+1} = \dot{y}_i$ (i = odd) are obtained from the same display indicator. If $f_{ci} > 0$ is the attention allocated to y_i for control purposes, then the constraints on f_{ci} are

$$\sum_{i=odd} f_{ci} = f_c \quad ; \quad f_{c,i+1} = f_{ci} \quad i = 1, 3, 5, \ldots \tag{30}$$

The pilot allocates his attention among displays, spending the larger f_{ci} on displays that are most useful for control.

With f_{ci} selected, the pilot-vehicle model yields predictions of the performance metric, P_c, Eq. (20). Using this prediction, we can study the tradeoffs between f_c and P_c for any given display system. Figure 2 is a typical performance/workload curve. It shows the performance attained for a given workload, as well as the workload required to obtain a given performance level. In Figure 2, the intersection of the line $P_c = P_{c,max}$ with the P_c versus f_c curve gives the minimum amount of control attention required, $f_{c,req}$, for the given system to meet P_c specifications. The difference between this amount of attention, and the total available for the entire task is the residual workload available for monitoring

$$f_{m,avail} = f_T - f_{c,req} \tag{31}$$

The process of comparing the candidate display systems is now clear. As an example, one may observe Figure 3. Clearly display system 1 is superior to display system 2, since less control workload (or <u>required</u> control attention) is needed to meet the required performance level. Moreover, more attention is available for monitoring duties when using display system 1.

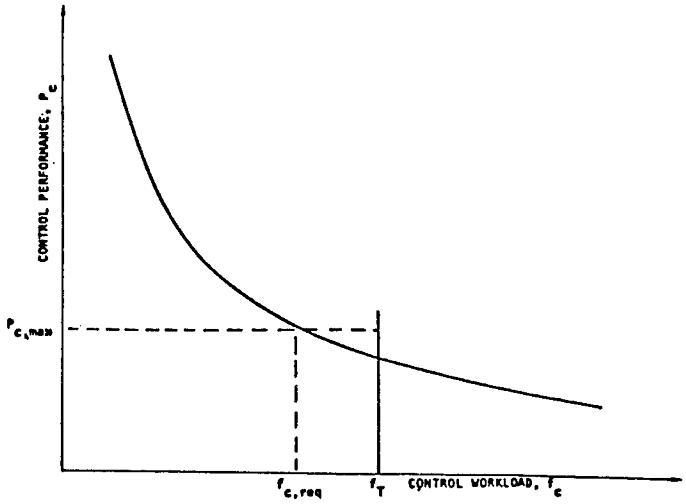


Figure 2. Conceptual Control Performance Versus Workload Curve.

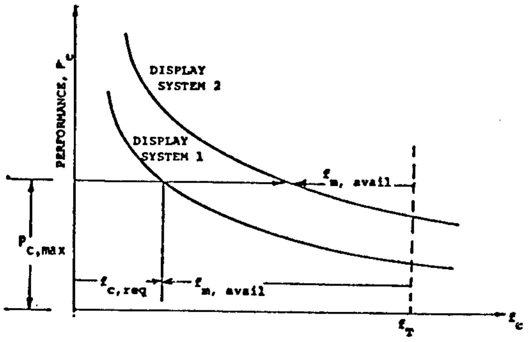


Figure 3. Guidelines for Evaluating Control/Display Designs.

In the terrain following task we compute the control performance metric P_c, Eq. (20), for the candidate systems given the constraints f_T = 0.2, 0.4, 0.6, 0.8 and 1.0. Also, the pertinent RMS flight path errors (e_{rms}) and g-levels (g_{rms}) at the same f_T values are obtained. These results are summarized in Table 2.

TABLE 2. PREDICTED CONTROL PERFORMANCE (P_C), FPE, (ft) AND g-STRESS (g's) RESULTS

Display System	f_T	0.2	0.4	0.6	0.8	1.0
Status	P_c	.84	.92	.95	.96	.97
	e_{rms}	28.6	22.6	20.4	19.0	18.0
	g_{rms}	.54	.50	.48	.47	.46
Tunnel	P_c	.87	.95	.97	.98	.99
	e_{rms}	26.2	20.4	18.0	16.6	15.8
	g_{rms}	.50	.46	.44	.42	.42
Predictor	P_c	.93	.98	.99	>.99	>.99
	e_{rms}	22.4	17.6	15.6	14.4	13.8
	g_{rms}	.44	.41	.40	.38	.39
Tunnel + Predictor	P_c	.94	.99	>.99	>.99	>.99
	e_{rms}	21.0	16.2	14.6	13.6	13.0
	g_{rms}	.45	.42	.41	.40	.39
Flight-Director	P_c	>.99	>.99	>.99	>.99	>.99
	e_{rms}	15.0	12.2	11.4	11.0	10.6
	g_{rms}	.41	.39	.38	.37	.37

Using these numerical results, the display systems attain the rank ordering as shown in Figure 4. Using Eq. (31) we are now able to compute the required control-attention, $f_{c,req}$ (workload), and the available monitoring attention, $f_{m,avail}$, for each of the candidate display systems. We assume f_T=.8. These results are summarized in Table 3.

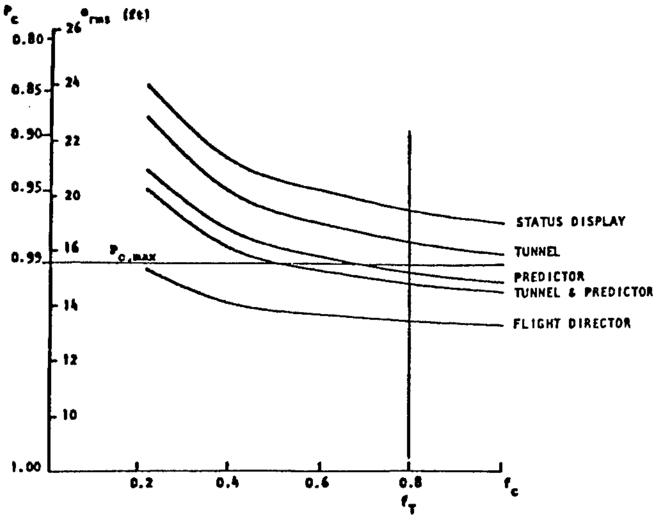


Figure 4. Display System Rank Ordering - Analytical Results.

TABLE 3. WORKLOAD AND MONITORING ATTENTION RESULTS

Display System	$f_{c,req}$	$f_{m,avail}$
Status	>>.8	0
Tunnel	>.8	0
Predictor	.64	.16
Tunnel + Predictor	.48	.32
Flight-Director	<.2	>.6

It is evident from these results that some form of synthesized display must be considered in the A10 aircraft, as the rudimentary status display yields an unsatisfactory control performance. It is also clear that the best system, in terms of reduced control workload, available monitoring attention, and lowest levels of g-stress, is the flight director display. In addition, the flight director's P_C-versus-f_C curve is almost flat (i.e. $dP_C/df_C \approx 0$) for $f_C \epsilon [0.4,0.8]$, which indicates that this design is robust with respect to external attentional demands that might be placed on the pilot. The tunnel, on the other hand, exhibits supprisingly poor results, as it ranks only fourth, after the predictor, and does not meet the performance criterion.

The predictor display requires a very high control attention, $f_{c,req}$=.64, and and the available monitoring attention, $f_{m,avail}$, is only .16. Although the design specifications are met, such a display system may not be acceptable, as it would be too sensitive to a possible degradation in control attention capacity. However, the integrated tunnel/predictor display system gives a satisfactory performance. Next we validate the modeling results experimentally.

EXPERIMENTAL VALIDATION

A primary objective of this study was the validation of the model-based display design procedure described previously. The validation phase consisted of fixed-base man-in-the-loop simulations of the A10 terrain following scenario for the four synthetic displays. The experiments, conducted largely independently of the analytic effort, were performed at the University of Connecticut.

DISPLAY FORMAT

A precursor to the experimental phase is the design of the display format, i.e. the details of the display panel layout. Clearly, this is largely an art, but can be guided by the results of the display element analysis with regard to threshold values and scale range. Four basic, or status, displays were used in the experiments in addition to the synthetic display. The basic displays were the following.

1. Error Indicator. This showed instantaneous error about the nominal terrain-following path. We used a vertical scale of ±50 ft range (recall 40 ft is the maximum design error). The distance between scale markings/divisions was chosen as 5 ft, which corresponds to a display threshold of ~2.5 ft.*

2. Pitch Indicator. A stylized aircraft pitch indicator was used to display $\theta(t)/q(t)$. A maximum range ±10° was allowed. Minimum scale marking was 2.5°.

3. g-Meter. Although vertical acceleration was not shown to be of significance as an observation, our analysis assumed that RMS g-level entered (subjectively) in the pilot's cost functional. Since our simulation was fixed-base, the only possible perception of g-level was via visual stimulus. Thus, the subjects were "aware" of their commanded accelerations.

4. Radar Altimeter. This display is essentially a duplication of the error information, i.e. the difference in altimeter reading from 200 ft is the error. It is included for those cases where the error indicator may be off-scale, $|e(t)|>50'$. In addition, any realistic display panel will likely contain this information.

Figure 5 shows the display panel layout that we used. The center screen area was set aside for the specific synthetic displays to be investigated. The display in the lower right corner is associated with a side monitoring task

---

*We generally assume that the display threshold is half the minimum scale marking or 0.05° visual arc, whichever is the larger. The display threshold should be less than the control indifference threshold, $y_{max}/8$, for a well-designed display.

which will not be addressed at this time. The entire display was presented to the subject on a VS60 graphics screen; the total display size was 14" x 12".

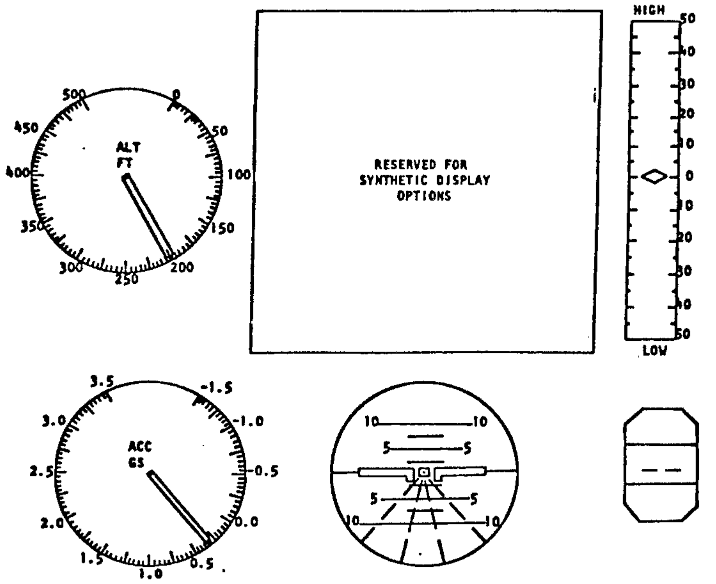


Figure 5. Basic Display Panel Layout for Terrain-Following Simulation.

The basic display format of Figure 5 was the same for all experiments. The only difference among the cases studied was the form of the synthetic display. These are now described.

Predictor Display

In this display we present the future terrain $\Pi(t+T)$ and extrapolated aircraft position $h(t)+T\dot{h}(t)$ where $T=4$ sec. The display format used is shown in Figure 6. Here the cross represents the aircraft flight vector. The "terrain-box" represents an 80'(H) x 100'(W) window centered on the terrain path at a distance $D_0=4xU_0\simeq1900'$ ahead of the aircraft. Thus, if the subject kept the cross within the box, the linear prediction of future error would be <40'. For convenience, the cross was fixed at the center of the synthetic display area, i.e., only the terrain-box moved.

Tunnel Display

The stylized tunnel display that was programmed for the experiments is shown in Figure 7. The "tunnel" consists of five "windows", separated in distance by 500'. Thus, with the extension lines, the tunnel presents the future path some 2500' - 3000' ahead of the aircraft. The tunnel is centered on the nominal terrain path and has longitudinal dimensions ±40' (to correspond with e_{max}) and lateral dimensions ±50'.

The tunnel (windows) are fixed in inertial space. Thus, as the aircraft "flies" forward, the tunnel windows move towards the observer. When the leading window reaches a minimal distance of 100' from the aircraft it disappears, and a "new" window appears at the tunnel's end. This gives the illusion of continual forward motion. The perspective view of the tunnel is along the aircraft's flight vector, $\gamma(t)$, i.e., the viewing axes are aircraft centered with the forward z-axis aligned with $\gamma(t)$.

In the present experiments, tunnel variations occur only in the longitudinal axis. However, the computer simulation can treat tunnel/terrain and aircraft motion in both longitudinal and lateral axes. A complete description of the computer simulation and software may be found in [12].

Integrated Tunnel and Predictor Display

This display format is essentially a combination of the two previous displays. The integration has been effected by adding an additional "window" to the tunnel display at a range D_O=1900' ahead of the aircraft. This window does not

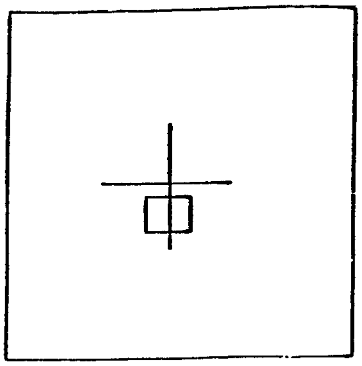


Figure 6. Predictor Display Symbology.

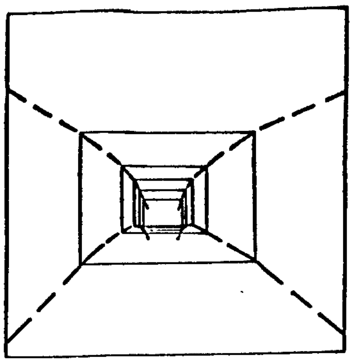


Figure 7. Tunnel Display Format.

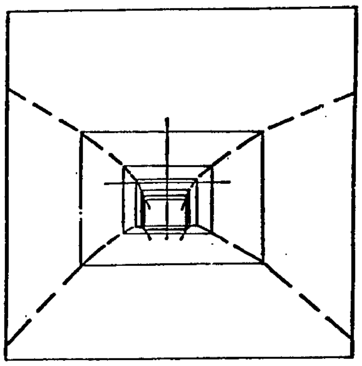


Figure 8. Integrated Tunnel and Predictor Display.

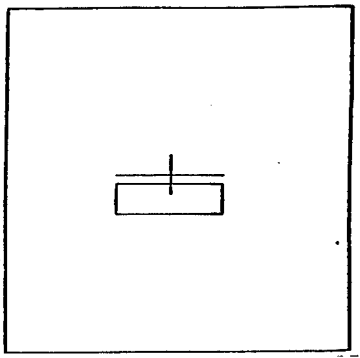


Figure 9. Flight Director Display Symbology.

move towards the observer, but the other windows pass through it. The display format is shown in Figure 8. In order that the predictor window be visually prominent it is shown brighter than the other elements that make up the tunnel display.

The velocity, or flight-path vector is superimposed on the tunnel as a cross. Since the tunnel view is centered on this vector, the cross remains stationary in the center of the viewing area. Of course, the tunnel (which is fixed in inertial space) curves up or down depending on the terrain and aircraft motion. We note that if the tunnel view were centered on the aircraft pitch angle, then the velocity vector projection would per force be different.

Flight Director Display

The flight director signal, $y_{FD}(t)$, as derived via the methodology described earlier is given by

$$\begin{aligned} y_{FD}(t) &= -.003\Pi(t) + .1277\dot{\Pi}(t) + .1095\ddot{\Pi}(t) \\ &+ .76\alpha(t) - .476q(t) - 1.245\theta(t) + .0594e(t) \qquad (32) \end{aligned}$$

This has been implemented as

$$y_{FD}(t)=c_1\Pi(t+\tilde{T})-c_2\Pi(t)+.76\alpha(t)-.476q(t)-1.245\theta(t)+.0594e(t) \qquad (33)$$

where $c_1 = .0745$, $c_2 = .0775$, $\tilde{T} = 1.715$ sec. In deriving Eq. (33) we use the approximation $\Pi(t+\tilde{T}) = \Pi(t) + \tilde{T}\dot{\Pi}(t) + (\tilde{T}^2/2)\ddot{\Pi}(t)$ and equate coefficients with Eq. (32). Note that c_1 may be assumed equal to c_2 with little or no observed effects.* In Eq. (33), $\Pi(t+\tilde{T})$ is the terrain path T sec. (or $\tilde{D}=\tilde{T}\cdot U_0 \simeq 800$ ft) ahead of the aircraft.

The flight director display is presented to the subject in the form of a compensatory tracking task, as shown in Figure 9. The cross in Figure 9 is stationary with respect to the viewing area; the signal that drives the box is given by Eq. (33). In order to distinguish this display from the predictor the box and cross sizing is different.

EXPERIMENTAL DESIGN

The fixed-base experiments were conducted using a PDP 11/60 to simulate the aircraft equations, terrain, and to update the displays. The display was presented on a VS60 graphics screen and refreshed 30 x per second. Pilot input was via a force stick controller. The specifics of the simulations are given in [6].

---

*This is convenient as only the path different $\Pi(t+T) - \Pi(t)$ is needed in the flight director signal.

Four subjects for the experiments were selected from the Air Force ROTC student body at the University of Connecticut. The display conditions were presented to them using a Latin square ordering to minimize any transition effects on averaged performance. A data trial lasted 130 sec., the last 0.06 x 2048 = 122.88 sec. of which was used as data. We recorded the 2048 samples of control input δ and error e for each trial. In addition, we computed and recorded RMS values of error, control, pitch and vertical acceleration for each run. Thus, we obtained a total of N = 8 x 4 = 32 trials for each display condition.

It should be noted that none of the subjects had flight experience; two had some fixed-base trainer experience. Thus, it is quite likely that the subjects were (uniformly) not expertly trained on the control task. However, it is quite likely that the relative differences in performance for different displays is not strongly dependent on absolute training level in the present task.

EXPERIMENTAL RESULTS

Table 4 gives the experimental results averaged across the four subjects. The averages were computed first for each subject and then across subjects to obtain the grand averages. The standard deviations in the experimental results, shown in parentheses in Table 4, are the averaged intra-subject variations and not the inter-subject variations. Thus, these numbers are indicative to run-to-run variability that might be associated with a single ("average") human.*

TABLE 4. AVERAGED EXPERIMENTAL RESULTS

Case	N	e_{RMS}	(ft)	$g_{RMS}{}^{-1}$	(g)
1	32	22.0	(3.4)	0.44	(0.068)
2	30	25.7	(4.5)	0.53	(0.078)
3	33	17.8	(2.2)	0.44	(0.084)
4	36	8.95	(1.9)	0.36	(0.054)

The results tabulated in Table 4 are quite interesting with regard to the underlying considerations in our display design technique. First note the rank-ordering of displays with respect to e_{RMS} performance. Here we see that the tunnel display (case 2) fares worst. The predictor display (case 1) fares slightly better than the tunnel alone. The combined tunnel plus predictor display (case 3) results in a meaningful performance improvement over that of (1) or (2). The flight director display (case 4) yields significantly better performance than any other display condition -- by a factor of 2. This is highly encouraging validation of our flight director design/synthesis procedure.

The rank-ordering of the different display configurations is in precise agreement with the analytical results of the display design methodology. While the

---

*If intersubject variability was included in the standard deviations, the values would increase by 20-100%.

absolute levels of control performance between experiment and model disagree slightly, relative performance levels agree well. This is demonstrated by comparing model predictions (at a fixed $f_c \simeq .6$) for cases 1-3 with the experimental results. The e_{RMS} experimental results for cases 1-3 are consistently higher than the model predictions. An explanation for this fact is that the OCM assumes a well-trained subject, whereas the actual subjects -- not being pilots -- were not fully trained with respect to the A10 dynamics. While it is possible to model this effect a posteriori in the OCM by increasing observation/motor noise and/or τ_N, this was not an objective of our efforts. On the other hand, the absolute performance levels for model and data in case 4 are in close agreement. The reason for this is that the "system" dynamics as perceived by the subject are similar to K/s. These dynamics are trivial to learn, so that training effects (after but a few trials) are inconsequential.

CONCLUSIONS

An analytic, pilot model-based, display design methodology has been applied to study workload and performance trade-offs in a high-speed, terrain-following task. The methodology combines pilot limitations, aircraft dynamics and performance requirements in order to determine the requisite information that must be supplied to the pilot.

Man-in-the-loop experiments that evaluated the performance of the four candidate display systems were conducted at the University of Connecticut. The objective of these experiments was to validate the overall display design procedure, including the flight director design synthesis process. Two positive conclusions resulted from this effort.

VALIDATION OF DISPLAY SYSTEM PERFORMANCE

The fixed-base experiments involved a terrain-following control task. (A secondary monitoring task was also included but is discussed elsewhere [6].) The experimental relative rank-ordering of the four display systems, based on control performance, was identical to that predicted analytically at the display element level. It suggests that a large number of potential display (or control augmentation) systems can be evaluated analytically at low expense, and then the most promising options can serve as the candidates for subsequent manned simulation. The time and cost savings of a model-based "front-end" to the complete design process can be substantial.

The spread in absolute levels of performance between the first three display systems and the flight director display was found to be much greater in the experiments than in the model predictions. Our explanation of this result is that the model assumes a well-trained pilot, whereas the subjects were not well-trained on A10 dynamics and so their performance was not at the model-predicted levels. On the other hand, the flight director essentially normalizes out the aircraft dynamics, rendering the control task much simpler and requiring virtually no learning. Here model and data absolute performance levels were commensurate.

VALIDATION OF FLIGHT DIRECTOR DESIGN PROCEDURE

The analytic technique for flight director synthesis is included in the display methodology at the information level. Here, we optimally synthesize or aggregate the information states into a single information variable that could be displayed to the pilot. The flight director signal is designed to relate to the pilot task objectives, i.e. to minimize his workload and/or improve his control performance, and to satisfy the pilot's desired goal of behaving approximately as a gain and time-delay. The man-in-the-loop simulations validated the superiority of the flight director display (over all others considered) with respect to control performance. The ability to analytically design a flight director that is in harmony with pilot control and information processing limitations is a major feature of the methodology. In many situations flight directors are "designed" via extensive simulations and tuning using a pilot in-the-loop. The analytic design, when used as prescribed, can shorten this experimental procedure to a great extent.

REFERENCES

1. Hoffman, W.C., D.L. Kleinman, and L.R. Young, "Display/Control Requirements for Automated VTOL Aircraft," NASA Contractor Report 158905, ASI-TR-76-39.

2. Curry, R.E., D.L. Kleinman, and W.C. Hoffman, "A Design Procedure for Control/Display Systems," Human Factors, 1977, 19(5), pp. 421-436.

3. Kleinman, D.L. and S. Baron, "Manned Vehicle Systems Analysis by Means of Modern Control Theory," NASA CR-1753, June 1971.

4. Kleinman, D.L., S. Baron, and W.H. Levison, "A Control Theoretic Approach to Manned Vehicle Systems Analysis," IEEE Trans. on Automatic Control, Col. AC-16, No. 6, December 1971.

5. Kleinman, D.L. and S. Baron," Analytic Evaluation of Display Requirements for Approach to Landing," NASA CR-1952, November 1971.

6. Korn, J. and D.L. Kleinman,"Advanced Cockpit Pilot Information Modeling," ALPHATECH TR-119, ALPHATECH, Inc., 3 New England Executive Park, Burlington, Massachusetts, 01803.

7. McRuer, D., I. Ashkenas, and D. Grahm, Aircraft Dynamics and Automatic Control, Princeton University Press, 1973.

8. Grunwald, A.J. and S.J. Merhav, "Vehicular Control by Visual Field Cues--Analytical Model and Experimental Validation," IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-6, No. 12, December 1976.

9. Grunwald, A.J. and S.J. Merhav, "Effectiveness of Basic Display Augmentation in Vehicular Control by Visual Field Cues," IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-8, No. 9, September 1978.

REFERENCES (continued)

10. Grunwald, A.J., "An Experimental Evaluation of an Electronic Perspective Tunnel Display for 3-D Helicopter Approach-to-Landing," TECHNION, Faculty of Aeronautical Engineering Report, TAE 383, October 1979.

11. Tomizuka, M., "Optimal Continuous Finite Preview Problem," IEEE Trans. on Automatic Control, June 1975.

12. Lovinsky, V.S. and D.L. Kleinman, "Software Documentation for TRNFLW," University of Connecticut, Department of EECS, Technical Report TR-81-10. June 1981.

APPENDIX

C_p AND D_p VALUES (EQ. 26)

T (sec)	$C_p(T)$	$D_p(T)$
1.5 (T*)	$\begin{bmatrix} -.007 & 1.43 & .83 & 12.25 & 0 & -12.25 & 1 \\ -.013 & .87 & .93 & -8.65 & 0 & -8.17 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ -.75 \end{bmatrix}$
4.0	$\begin{bmatrix} -.09 & 3.03 & 3.43 & 32.66 & 0 & -32.66 & 1 \\ -.050 & .39 & .95 & -36.68 & 0 & -8.17 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ -2.0 \end{bmatrix}$

DEVELOPMENT OF A PILOT MODEL FOR HELICOPTER VISUAL FLIGHT TASK SEGMENTS*

A. V. PHATAK, M. S. KARMALI
Analytical Mechanics Associates
Mountain View, California

and

E. J. Hartzell
Aeromechanics Laboratory, USARTL
Moffett Field, California

SUMMARY

This paper addresses the problems associated with developing an analytical representation for the human pilot in helicopter visual flight task segments. A two-level hierarchical model structure with elements corresponding to the autonomous information processing and control tasks and higher level decision-making functions is proposed. The utility of this modeling framework for understanding or interpreting pilot response behavior is discussed with reference to the visual approach to a hover task. An information-theoretic approach for rank ordering the visual cues according to information content is developed, and applied to the austere helipad scenario.

1. INTRODUCTION

An important problem in helicopter mission task analysis is the assessment or prediction of task performance and concomitant pilot workload in accomplishing the mission objectives. Task performance measures are relatively easier to quantify and may be inferred through a human factors analysis of the piloting task. On the other hand, finding an analytical expression for the degree of workload experienced by the pilot in performing a given flight task is not so obvious. This situation persists in spite of the voluminous amount of material published on the subject of workload [1] in the human factors literature. As a result, the prevalent approach to pilot performance and workload evaluation is empirical and is based upon the interpretation of pilot opinion ratings [2] which must be obtained through extensive piloted simulations and/or flight tests. However, such an approach is strictly experimental

---

* This effort was performed under Contract No. DAAK51-81-C-0004 from the Applied Technology Laboratory, USARTL (AVRADCOM) For Eustis, Virginia.

and provides little insight into the identification of the underlying factors influencing task performance, pilot workload, and ultimately pilot opinion.

A fundamental axiom in pilot-vehicle systems analysis states that the pilot's acceptance of an aircraft for a particular mission under a given set of operating conditions involves a tradeoff between two key factors: 1) task performance as perceived by the human pilot, and 2) the degree of workload imposed upon the pilot in achieving this performance level. The pilot in a realistic flight control task must use the information available from the various sensory modalities to assess task performance and generate the control strategy appropriate for accomplishing the desired task objectives. Hence, for a given acceptance level of task performance, pilot workload may be related to the information processing, decision-making and control requirements imposed upon the human pilot in satisfactorily accomplishing the task objectives.

Thus, an analytical methodology based upon the use of an appropriate pilot model structure is needed. Existing pilot models; namely, the describing function model [3] and the optimal control model [4], are too structured for describing pilot behavior in visual flight control tasks encountered in terrain flight missions. These models are primarily useful for describing human response during flight tasks under instrument-meteorological conditions. The unique features of helicopter-visual flight task segments, in particular, the availability of the extra-cockpit visual scene, preclude the use of existing pilot model based methodologies. A "bottom up" or task motivated approach to pilot model formulation is needed.

In an effort to remedy this situation, the U.S. Army Aeromechanics Laboratory at the NASA Ames Research Center has initiated a comprehensive, long-term program committed to the development of a cohesive analytical framework for improved pilot-vehicle performance and workload analyses. This study was motivated by the desire to develop an integrated quantitative approach to the design of the pilot-helicopter interface, that will allow full exploitation of both pilot capabilities and advanced technological developments.

This paper is organized into four sections. Section 2 presents a pilot model structure that is suitable for describing pilot-behavior in a wide variety of flight tasks. The utility of the proposed model structure is demonstrated in Section 3 and the Appendix by application to a typical mission segment--namely, the visual approach to a hover over a prescribed landing pad. Problems associated with formulating an analytical representation of the pilot's information processing, decision-making and control strategies are discussed. Conclusions and recommendations for further work are given in Section 4. Details may be found in Reference 5.

2. PILOT MODEL FORMULATION

The helicopter pilot involved in a civilian or military mission is required to perform a wide range of tasks. The specific nature of these tasks is dictated by a number of factors, the most important among these being: (1) the mission objectives, (2) the operating environment and scenario, (3) the helicopter type, configuration and dynamic characteristics, and (4) the sophistication of the groundbased and airborne navigation and guidance system.

The principal objective of this paper is to present a conceptual framework for analyzing human information processing, decision making and control behavior in flight segments representative of typical civilian and military helicopter missions. Therefore, this study is limited to an investigation of key piloting task segments such as approach to a hover, hover (or stationkeeping) and nap-of-the-earth (NOE) using extra cockpit visual cues as the primary source of navigation information. The pilot model structure formulated here is based upon the premise that, although the various missions and flight segments appear to be different at the outset, there exists a substantial degree of commonality in the nature of the task functions and requirements imposed on the pilot/crew.

Fundamentally, the process of flying consists of performing a hierarchy of tasks corresponding to increasing levels of cognitive involvement. As a minimum, piloting tasks may be classified into two hierarchical categories: (1) lower level autonomous information processing and control tasks, and (2) higher level decision-making tasks. A hierarchical model structure for the pilot based upon the two level task decomposition is shown in Figure 1.

The autonomous information processing and control task consists of a sequence of four operations; a) cue selection, b) cue information processing, c) performance evaluation, and d) control law implementation. The pilot has access to a variety of cues from different sensory modalities--namely, visual, vestibular (motion), proprioceptive and accoustic. This discussion, however, is limited to a consideration of the visual cues as the primary sensory information available to the pilot. Generalization of the concepts to an array of sensory cues appears feasible and may be carried out in subsequent studies.

The extra cockpit visual scene provides the pilot with a rich array of visual cues which contain information about the aircraft situation with respect to the outside world. Since the pilot does not have unlimited data processing capability, he must choose the necessary cues containing the most information content about the aircraft situation for estimating the state of the aircraft. The cue selection and information processing elements perform the navigation task of determining the aircraft position and orientation from the available visual cues.

The performance evaluation element compares the estimated aircraft state with some desired reference and determines whether or not a control response is warranted. A simple deadzone nonlinearity represents one plausible form for the performance evaluation logic element. In this case, control action is dictated only when the estimated aircraft state deviates from the desired value beyond some prescribed threshold level. The control law element performs the task of implementing control actions that tend to drive the aircraft state to the desired reference situation.

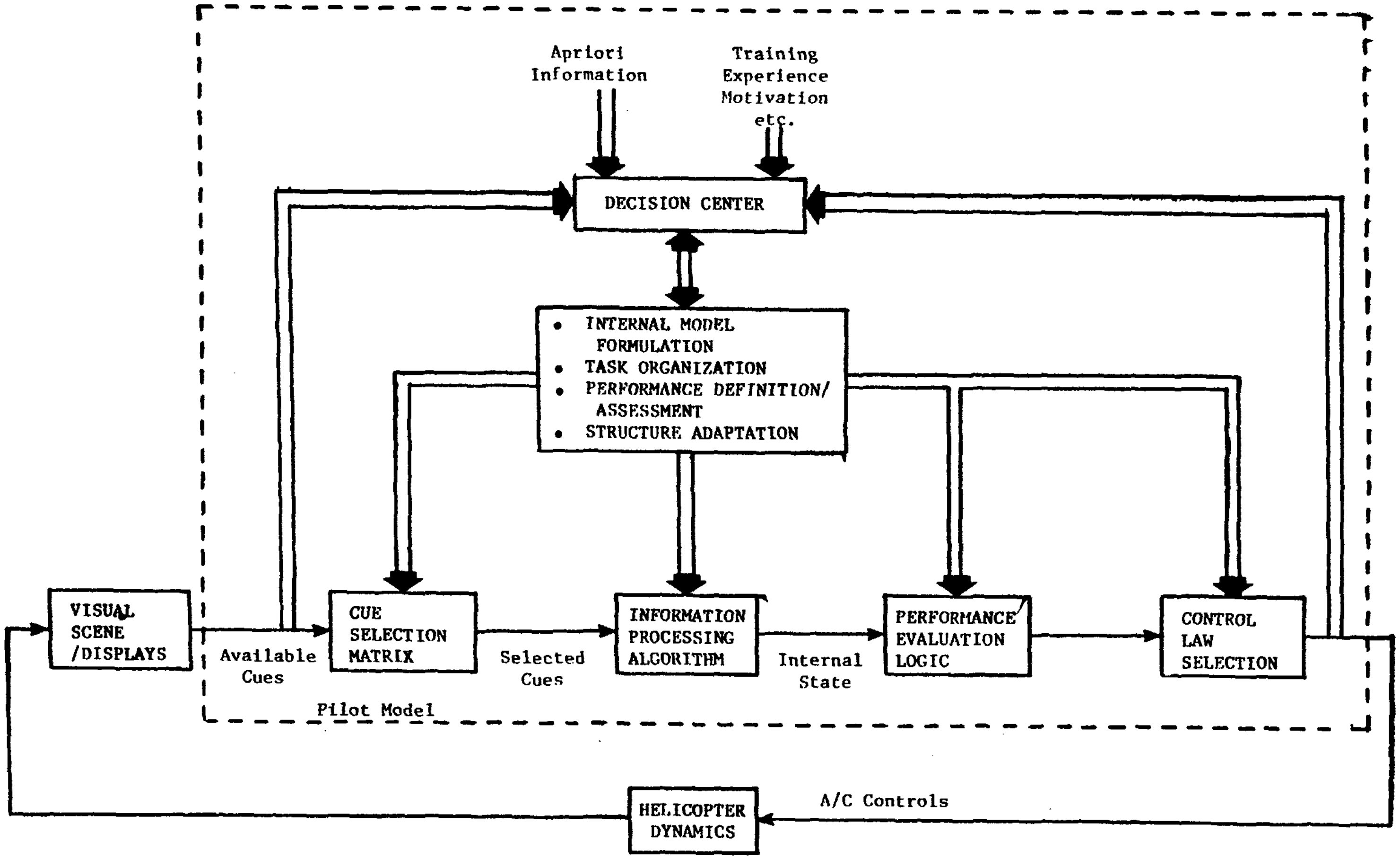


Figure 1. Hierarchical Pilot Model Structure

The decision center performs the higher level cognitive functions consisting of a) internal model formulation, b) task organization, c) performance definition and assessment, and d) structure adaptation.

The term "internal model" refers to the form of the pilot's perception of his own (aircraft) situation with respect to the outside world. This is equivalent to making assumptions about a frame of reference that may be used by a pilot in performing the flying tasks. The standard approach is based upon physical considerations and assumes that the pilot adopts an earth-fixed (cartesian, spherical or cylindrical) reference frame with it's origin at some meaningful point on the earth (e.g., helipad center). This approach leads to the definition of the perceived aircraft situation in terms of the familiar position (x,y,z) and orientation (ψ,θ,ϕ) state vector representation. However, it is quite likely that a pilot may use an entirely different precept based upon phenomenological rather than physical considerations. An example, is the use of geometric features in the visual scene as internal states for accomplishing a given flight task.

The process of task organization consists of defining the particular task, as perceived by the pilot, into an analytical framework that is consistent with the pilot's internal model or representation of the flight scenario. The mission objectives must be translated into a desired mission profile consisting of a concatenation of the individual phases and task segments. Eventually, this must lead to a definition of an acceptable or desired performance envelope for the instantaneous aircraft state as a function of some independent variable such as time or range-to-go. In an earth referenced internal frame of reference, this usually implies the specification of a desired, range dependent, reference flight trajectory and speed profile for the aircraft.

The function of the structure adaptation element in the decision center is to choose the structure and parameters for the cue selection matrix, information processing algorithm, the performance evaluation logic and the control strategy algorithm that are most suited for the particular task segment and operating scenario. This function reflects the well-known capability of human pilots to fly intelligently by adapting their information processing and control strategies to varying internal or external conditions.

The proposed model structure is extremely versatile and allows for considerable flexibility in selecting the mathematical representation for each of the individual elements. In the absence of certain knowledge or conviction regarding the actual operations performed by a pilot, the only rational recourse is to hypothesize plausible mechanisms to represent the various piloting processes.

The utility of this hierarchical model structure for understanding and interpreting pilot response behavior in realistic flight control tasks is illustrated in the next section with reference to the visual approach to a hover task.

3. VISUAL APPROACH TO A HOVER TASK

The ability of the helicopter pilot to fly a visual approach to hover over a prescribed landing zone is a well-known and routinely accepted skill. What is not understood, however, is the process by which the human pilot is able to perform this difficult task. Based upon the pilot model structure formulated in the previous section, the human pilot may be described as performing the following three functional tasks:

1) Acquire and process the necessary visual cues to estimate the position and orientation of the aircraft with respect to the prescribed landing pad,

2) Assess the estimated aircraft situation or state by comparing it with the desired objectives, and decide whether a corrective control action is warranted, and

3) Apply a corrective control action so as to bring the aircraft to an acceptable or desired state.

This section presents the plausible rationale and schemes that may be used by a pilot in accomplishing the above three functional tasks.

Section 3.1 describes the processes that may be involved in the pilot's selection of the visual cues inherent in the visual scene. An estimation-theoretic rationale for cue selection that is based upon quantifying and rank-ordering the information content in the visual cues is proposed and described in the Appendix. The usefulness of this approach is demonstrated by application to a generic cue set that is representative of the helipad approach scenario.

Having discussed an approach for selecting the visual cues, the next step is to describe the pilot's use of such cues during the helicopter visual approach and hover tasks. Section 3.2, discusses this issue from the phenomenological as well as physical viewpoints. Problems associated with interpreting the pilot's information processing, decision-making and control functions into a mathematical closed loop simulation of the visual approach to a hover task are discussed.

Finally, Section 3.3, presents some thoughts on factors contributing to pilot workload within the proposed pilot modeling framework.

3.1 Selection of the Visual Cues

The extra-cockpit visual scene provides the pilot with a wide variety of cues which he may use for aircraft navigation. Just what these cues are, and how they may be used by the human pilots are questions that have been a subject of intense research and debate among scientists over several decades. There are at least two distinct approaches towards determining the answers to these questions--the physical and the phenomenological. The physical approach is based upon treating the visual scene as an image (provided by the retina of the eye) which contains information about the aircraft state relative to the earth. Information theoretic methods can then be used to determine and rank order the information content in the visual cues. In contrast, the phenomenological approach starts with the notion that the human pilot alone is best qualified to identify visual cues in the extra cockpit scene and define ways to use them. Neither of these two approaches is without drawbacks. The physical approach, by definition, depends entirely on a deductive method and may lead to conjectures that appear entirely speculative and unreasonable from the human information processing viewpoint. The phenomenological approach which depends upon the use of pilot questionnaires is criticized for just the opposite reasons--namely, the over-emphasis on pilot interpretation of the visual cues which presumes that a pilot is consciously aware of what he truly does, and furthermore, that he can articulate this knowledge in specific terms. Therefore, both of the above two approaches were followed in this study. A pilot questionnaire designed to elicit answers to key questions was prepared. Results obtained from this questionnaire study will be reported separately. The discussion in this paper is limited to an analysis of the visual cues from an estimation/information-theoretic viewpoint.

The literature on visual cues is extensive [6-10]. However, it is generally agreed that monocular cues are more than adequate for visual piloting tasks [10]. The monocular cues commonly listed [11] are: (1) linear perspective, (2) relative motion or movement parallax, (3) apparent versus real size and shape, (4) interposition, (5) relative contrast and brightness, (6) aerial perspective, (7) texture, and (8) accommodation.

Examination of the above list of cues indicates that the majority of the cues are geometric in nature and can be described using methods of perspective geometry. The geometric visual cues available to the pilot during an approach to a helipad are discussed next.

3.1.1 Geometric Visual Cues

The visual scene as sensed by the retina is a two-dimensioned (2-D) image of the three-dimensional (3-D) world. As such, this 2-D image is ambiguous and does not provide a unique characterization of the actual 3-D visual scene. This is because the projection or mapping of a 3-D object onto a plane is many-to-one. The ability of humans to resolve this ambiguity by associating a unique external 3-D representation is based upon prior knowledge, learning, and interpretation by the central nervous system. A perspective static view of the approach to a hover task scenario is shown in Figure 2. This discussion is limited to helicopter motion in the vertical-longitudinal axes. The coordinates (x,0,h) represent the instantaneous position of the position of the helicopter in a helipad reference frame with its origin at the center of the pad. Figure 3 shows the static forward view out of the cockpit as seen by the pilot. For the purpose of this discussion, the pilot's line-of-sight is assumed to be level and aligned with the helipad's longitudinal axis. Thus, for a helipad ABEC of length 2L and width 2D, the projected forward image on a canopy plane placed at a normal distance c from the pilot's eye corresponds to the trapezoid A'B'E'C'. Similarly, any horizontal line element PQ on the helipad maps into a corresponding line element P'Q' in the projected image. The three variables ℓ_{x_0}, w_{x_0} and η shown in Figure 3 characterize the projected image of the helipad. For the scenario described by Figure 2 and 3, ℓ_{x_0}, w_{x_0} and η can be shown to be nonlinear functions of the aircraft coordinates x_0 and h. Thus,

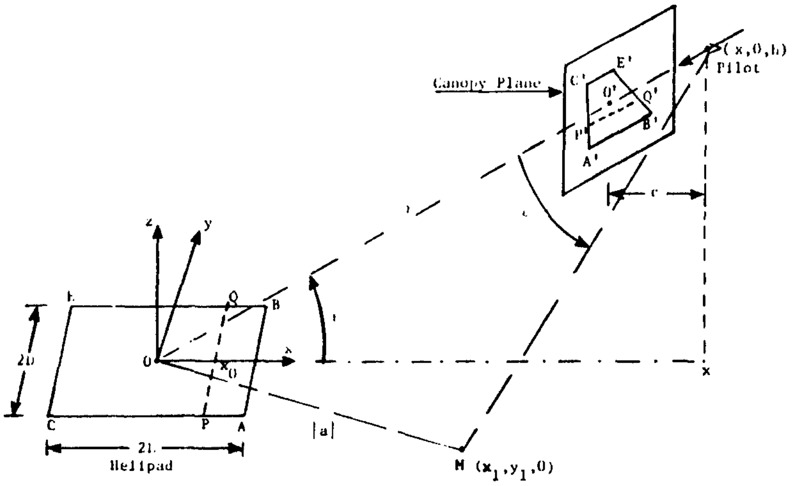


Figure 2. Helipad Approach Scenario

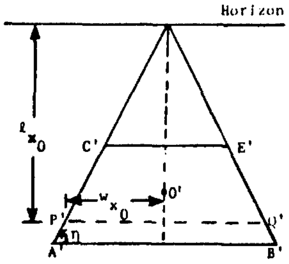


Figure 3. Perceived Extra-Cockpit Visual Scene

$$\ell_{x_0} = \frac{ch}{x-x_0} \qquad (1)$$

$$w_{x_0} = \frac{cD}{x-x_0} \tag{2}$$

$$\eta = \tan^{-1} \frac{h}{D} \tag{3}$$

The static perspective view of the helipad is a source of a number of visual cues to the pilot. Plausible candidates include (1) perceived length of the helipad, (2) perceived widths of the helipad at the far end, near end and center, (3) perceived inclination of the helipad edges to the horizontal, and (4) perceived depressions of the far end, near end and center of the helipad with respect to the horizon. The cues defined above only represent one plausible set. Other reasonable static cues involving ratios or sums of perceived lengths and angles may quite well be used by the human pilot. Furthermore, the framework of perspective geometry may also be used to formulate visual motion cues [12] (i.e., motion cues obtained from the dynamic frame-to-frame changes in the perceived visual scene) and texture cues.

Based upon the above discussion, the pilot has a multitude of visual cues available to him for performing the navigation, guidance and control functions. For the scenario described in Figures 2 and 3, only two independent measurements or cues are necessary for estimating instantaneous helicopter longitudinal range x and altitude h. However, the pilot is not limited to using only two cues. Just how the pilot chooses a given set from the vast number of plausible cues is not yet understood. The approach taken here is to assume that the pilot chooses the visual cues which contain and provide the most amount of information for the navigation task--namely, the task of estimating vehicle position and orientation with respect to the landing pad. The Appendix presents an analytical method for determining and ranking the information content in the visual cues. The discussion is limited to geometric visual cues in the visual approach to a hover task; however, extension to any other set of cues is straightforward. An information theoretic analysis of the visual cues defined by Eq. (1)-(3) above shows that the amount of information contained in these cues grows with increasing values of x_0 and D, or equivalently an enlarging field-of-view in the elevation and azimuth planes, respectively.

3.2 Task Utilization of the Visual Clues

The previous section and the Appendix describe a plausible rationale that a pilot may use in selecting a set of visual cues from the extra-cockpit scene. The material in this section attempts a similar formalization of the functional process involved in the pilot's utilization of the visual cues in the approach to a hover task. A first step in this process is to try and understnd the pilot's subjective interpretation or perception of the guidance requirements for the given task as described next.

3.2.1 Guidance Commands

A reasonable way to start is by defining the task objective. For a visual approach to a hover task the principal objective is to bring the helicopter from some initial cruise altitude and velocity to a hover over the prescribed landing pad. In order to accomplish this goal, the pilot must apply the necessary control inputs (primarily longitudinal cyclic and collective for the straight-in profile) to the helicopter to fly a desired descending and decelerating approach profile. This implies that the pilot has developed some explicit and/or implicit notion of a desired approach profile or corridor during the approach to a hover task.

The existence of such preferred visual approach profiles, or corridors, is clearly demonstrated by the results of a flight experiment, by Moen, et al. [13], showing definite patterns or characteristic shapes in the altitude, ground-speed, and deceleration profiles of visual approaches for helicopters. Two key features are apparent: (1) the altitude versus range profiles resemble a typical three segment approach trajectory consisting of a) a constant altitude segment, b) a constant flight path angle ($6.5^o \sim 12.5^o$) approach and c) a final constant altitude segment prior to hover, and (2) the ground speed versus range profiles correspond to a constant g (≈ .03g) deceleration, except for the last 1000 ~ 1500 feet prior to hover when the speed is suddenly reduced by an almost exponential increase in deceleration. Furthermore, definite trends in these patterns with initial altitude and speed are observed. These results reinforce the belief that a trained pilot deliberately chooses to fly a desired approach profile under a given set of operating and environmental conditions.

Just how the pilot performs the guidance task is a matter of conjecture. Formulation of a plausible guidance strategy depends upon what one assumes for the pilot's internal model or representation of the outside world. The approach taken here is based upon the assumption that the pilot perceives the outside environment in terms of the helicopter position and orientation with respect to a helipad centered reference frame (i.e., cartesian, cylindrical, or spherical coordinates). Thus for a straight-in approach (along the x axis of the helipad as shown in Figure 2), the helicopter state or position coordinates may be represented as:

$$\underline{x} = (x,h)^T \text{ in cartesian coordinates, or} \qquad (4)$$

$$\underline{x} = (\gamma,\ h)^T \text{ or } (x,\gamma)^T \text{ in cylindrical coordinates} \qquad (5)$$

where $$\gamma = \tan^{-1} \frac{h}{x} \qquad (6)$$

is the flight path angle. Therefore the desired altitude and ground speed profiles may be expressed as:

$$\left.\begin{array}{l} h_c = h_c(x), \text{ or} \\ \gamma_c = \gamma_c(x) \end{array}\right\} \quad : \text{ Altitude Command} \tag{7}$$

and

$$v_c \equiv \dot{x}_c = v_c(x) \quad : \text{ Speed Command} \tag{8}$$

Note that the formulation of the guidance commands given by Eq. (4)-(8), is motivated by physical concepts commonly used in autopilot design. However, it is likely that a pilot may use an entirely different precept based upon phenomenological rather than physical considerations. An example, is the use of feature patterns in the visual scene as internal states for accomplishing a given flight task. The key lies (1) in recognizing visual cues or feature patterns which exhibit invariance properties for a desired flight profile, and (2) in using these identified cues for structuring the guidance commands. There have been several studies in the past [14-23], directed towards the subject of visual cues during the approach and landing task. These studies show that the pilot may be using a number of invariance patterns in the visual scene. Examples for a constant glide slope visual approach are (1) the constant depression of the intended touchdown point below the horizon, and (2) the stationarity of the aim point (focus of expansion) with respect to other objects in the visual scene. Thus, for a constant desired flight path angle γ_c, the guidance command

$$\gamma_c = \text{constant (e.g., } 6.5^o \sim 12.5^o) \tag{9}$$

may be equivalently described as

$$\ell_0 = c \tan \gamma_c \equiv c \frac{h}{x} = \text{constant} \tag{10}$$

where ℓ_0 is the depression of the center of the helipad (intended aim point) below the horizon. Note that in the absence of a horizon, an alternate invariance pattern for constant γ_c is

$$\frac{\ell_L - \ell_0}{2 w_L} \equiv \frac{L}{2D} \tan \gamma_c = \frac{L}{2D} \frac{h}{x} = \text{constant} \tag{11}$$

where w_L, ℓ_L and ℓ_0 are defined by Eq. (1)-(3). Equation (11) states that a constant glideslope approach to the center of the helipad can be flown by maintaining the ratio of the perceived distance of the helipad center from the near end of the helipad to the perceived near end width at a constant value.

Equations (10)-(11) show that a constant glide slope approach can be flown without using cues that permit an explicit estimation of the ground range x or altitude h. However, these constant glideslope conditions only hold for a portion or segment of the overall approach trajectory. Therefore, the helicopter pilot must use additional cues that provide information for range and altitude estimation.

The above discussion indicates that several hypotheses can be advanced to describe the guidance command structures that may be adopted by the humam pilot. The following paragraphs discuss the issues involved in integration of such guidance schemes with the navigation, performance assessment, and control functions of the helicopter pilot in the visual approach to a hover task.

3.2.2 Closed-Loop Simulation of the Visual Approach to a Hover Task

Before one can proceed with a closed loop simulation of the visual approach to a hover task, it is necessary to have a detailed mathematical representation for each of the elements in the pilot model structure. The preceding material presented some plausible schemes for selecting the visual cues and for defining the guidance laws. This section discusses the formulation of algorithms for the cue processing (i.e., state estimation), performance evaluation and control elements in the hierarchical pilot model structure.

As stated earlier, the pilot model structure of Figure 1 is extremely versatile and allows for considerable flexibility in choosing the form of the individual functional elements. Thus, the describing function [3] and optimal control model [4] structures represent two specific formalizations within the hierarchical model structure. The describing function lumps all aspects of human information processing and control into a "black box" input-output representation. At best, higher level cognitive functions are accounted for by allowing for an adaptation of the describing function parameters (gain, lead, lag, time delay, etc.) with changes in the operating flight conditions. In contrast, the optimal control model formulation represents an attempt towards developing a model structure that is homomorphic to known psycho-physical aspects of human response behavior. Specifically, the model consists of three elements in tandem--a Kalman filter or state estimator, an optimal predictor, and an optimal linear quadratic controller. The optimal control model structure is normative; that is, the model describes or predicts what the human should do as opposed to what he actually does do. Furthermore, according to the optimal control model, the structures for the pilot's cue selection, information processing, performance evaluation and control are predefined within the linear, quadratic, and Gaussian (LQG), estimation and control theoretic framework. However, even a cursory review of the visual approach to a hover task is sufficient to indicate that such a model in its standard form may not be suitable, however attractive its use may appear, for describing the human pilot in this situation.

During visual meterological conditions (VMC), the out-of-the-window visual scene and the standard cockpit instruments are considered to be the two primary sources of information available to the pilot. Hence, it is reasonable to assume that the quality of the available cues in the visual scene will have a direct impact on the pilot's selection of the visual cues (including those from cockpit instruments) and their subsequent utilization in the navigation (i.e., state estimation), guidance (i.e., desired approach profiles), performance evaluation, and control tasks. Consequently, the shape of the desired altitude and velocity guidance profiles would be expected to adapt to the given operational scenario and environmental conditions. In this context, the nominal visual approach profiles for altitude and ground speed versus range, [Figures 5 and 8, respectively, in reference 13], were obtained at NASA Wallops Flight Center under normal weather conditions using helicopters equipped with standard cockpit instruments. It is important to note that the nominal flight profiles are not likely to remain unchanged under austere environmental conditions or in situations where the pilot is denied information from the cockpit instruments.

The helipad landing scenario described by Figure 2 is definitely austere, and was deliberately chosen to permit an understanding of the visual cues from first principles. Unfortunately, there is practically no data available for this scenario that could be used for objectively defining the guidance profiles and the corresponding navigation and control algorithms adopted by the human pilot. In the absence of data, the only option available is to proceed with the simulation using plausible mathematical representations for the various modules in the human pilot model. A matrix of off-line computer simulation tests may be defined corresponding to varying degrees of information processing and control law sophistication. Thus, for example, the navigation algorithm for estimating the aircraft state from the given visual cues can be as simple as a deterministic inversion of the nonlinear mapping describing the visual cues or involve the implementation of sophisticated model-based filters (i.e., α-β filter, complementary filter of Kalman filter in that order). Similarly, the performance evaluation and control module structures can range from the classical or optimal (LQG) feed-forward/feedback control laws to an alternative control concept, recently introduced by Rault and Richalet [24,25], termed "model predictive heuristic control" [24], "scenario predictive control" [25], "model algorithmic control" [26], or "output predictive algorithmic control" [27].

The alternate control technique is radically different in concept from the existing methods (i.e., classical control, LQG control, etc.) which rely on the explicit feedback of the instantaneous state estimate to achieve the desired objectives. Instead, the technique is based upon the computation of the control input, every Δt seconds (not necessarily constant), so as to minimize the predicted error over some future "horizon of prediction" between the actual and desired output profiles (e.g., altitude and range-rate profiles). This control technique has been successfully applied to a number of practical aircraft problems [26-27]. This approach appears to provide the most natural framework for describing human control strategy in situations such as the visual approach to a hover task.

The above discussion points out the wide range of plausible information processing and control algorithms, hence, the enormous dimensionality of the resulting closed-loop simulation test matrix that must be investigated.

3.3 Interpretation of Pilot Workload

Pilot workload is usually defined as an aggregate scalar measure of the degree of effort or difficulty experienced by the pilot in accomplishing a task at a given level of performance. However, the difficulty experienced by a pilot depends upon the nature and complexity of the functional requirements imposed upon the pilot in performing the task. Therefore, it is preferable to conceive of workload as a vector metric with elements representing the contributions of the pilot's information processing, decision-making and control tasks. Then, aggregate workload can be defined as some linear or nonlinear function of the vector components. Three approaches for quantifying pilot workload have been presented in the literature. They are: (1) subjective workload or demand ratings, (2) secondary task performance measures of spare mental capacity, and (3) physiological indicators such as galvanic skin resistance (GSR), pupil dilation, and heart rate. A comprehensive survey of the workload literature, including a useful taxonomy of the workload measures and assessment methods, may be found in reference [1]. These techniques, however, do not represent valid analytical measures for pilot workload; therefore, methods based upon an analytical representation or model of the human pilot's information processing, decision-making and control strategies have been proposed. Thus, the describing function model based workload measure is defined to be some nonlinear function of the model parameters associated with task difficulty--namely, the lead parameters (time constants fo the first order zeros) reflecting prediction requirements and the time delays signifying task urgency [28]. Similarly, the optimal control model based measure of workload is assumed to be proportional to the fraction of total attention capacity that must be devoted to the task in order to maintain a given level of performance [29]. Clearly, the validity of such measures is only as good as the veracity of the pilot model structures on which they depend. As discussed earlier, the applicability of the existing pilot models to the visual approach to a hover task is questionable, and a "bottom-up" or task motivated approach to pilot modeling and performance/workload assessment is recommended.

Analytical measure reflecting the degree of effort of difficulty experienced by the pilot performing the task are required. In the visual approach to a hover, the visual cues provided by the visual scene and standard cockpit instruments are the pilot's primary source of position and orientation information. Therefore, pilot workload would be expected to increase with degradation in the quality (i.e., information content) of the visual scene. Pilot workload for a given scenario and acceptable level of task performance must depend upon (1) the complexity or sophistication of the information processing, decision-making and control algorithms, and (2) the amount of control effort required. The control effort required can be quantified easily in terms of its variance and frequency content (i.e., spectral density characteristics). However, defining analytical measures for information processing, decision making, and control law complexity, is not so obvious. This is especially true when dealing with human cognitive functions and abilities as opposed to computer capabilities. A case in point is the phenomenal human capability for effortlessly analyzing and processing natural cues as provided by the human senses. Here the task of spatial orientation and performance assessment becomes easier as the quality of the visual scene improves, and vice-versa. This observation is counter-intuitive from the computer processing viewpoint. The key lies in recognizing that the human comes equipped with a built-in cognitive capability that allows the parallel processing of natural cues. Hence, information processing complexity, as perceived by the pilot, decreases with increasing number of visual cues as long as these cues are integrated and conform to natural (i.e., compatible with human sensory apparatus) format.

Unfortunately, making similar observations and comments about the pilot's control strategy is not so straight-forward. For example, consider two control strategies based upon classical (output feedback) and linear quadratic control which are designed to yield the same performance and require identical amounts of control effort. Clearly, from a computer implementation viewpoint, the linear quadratic controller is much more complex than a classical controller since the former requires a solution to a matrix Riccati equation for determining the control gains. This observation also holds from the piloting viewpoint unless one assumes that the human also comes equipped with a built-in processor for solving Riccati equations. Obviously more experimental work aimed towards identifying pilot's preferences in control strategy selection is needed.

The above discussion points out the difficulties involved in formulating analytical measures for pilot workload. However, much understanding about the piloting task requirements can be gained by performing an analysis of the tradeoff between the information content in the visual cues and required control effort, for a given level of task performance.

4. CONCLUSIONS AND RECOMMENDATIONS

A systematic pilot model-based framework for analyzing and interpreting pilot response behavior in visual flight tasks is proposed. A generic pilot model structure incorporating the known human functions of information acquisition/processing, decision making and control is formulated. The utility of this model-based approach towards understanding the piloting task is discussed with references to the helicopter visual approach to a hover scenario.

The extra-cockpit visual scene is assumed to be the pilot's primary source of information. The helicopter pilot must use this information to perform the required navigation, guidance and control tasks. Therefore, the pilot workload in accomplishing the task at a satisfactory performance level must be intimately dependent upon the quality of the visual scene from the information viewpoint. To understand this relationship, it is first necessary to have some quantitative measures of information content in visual cues. Hence, an estimation-theoretic approach towards quantifying and rank-ordering the information contained in the visual cues available from the extra-cockpit scene is described. The basic concepts are elucidated by analyzing and ranking the geometric visual cues to the pilot during an approach to a helipad.

The weakest link in applying a model-based approach lies in not being able to define what a pilot actually does with the information provided by the visual cues. In other words, no firm data base of rationale exists for selecting specific algorithms to describe the information processing, decision, and control elements in the proposed model structure. However, plausible mechanisms for describing the pilot's utilization of the visual cues in performing the flying task are discussed.

Pilot workload is defined as a vector quantity with elements corresponding to the information processing, decision-making and control components. An interpretation of workload in terms of the required model complexity and control effort for accomplishing a task is presented.

In summary, a cohesive structure for the analysis and interpretation of human pilot behavior in realistic flight scenarios has been developed. However, an experimental data-base is needed to develop analytical formulations for the pilot's information processing, decision-making and control algorithms. Two types of piloted simulation experiments are recommended: 1) Laboratory fixed base simulation studies, and 2) Flight test experiments. The purpose of the labortory tests would be to determine some fundamental characteristics of human perception in the context of visual flight tasks. Flight tests, on the other hand, are needed to provide a realistic data base for interpreting the effects of the visual scene content on the piloting task.

APPENDIX

Information Theoretic Analysis of the Visual Cues

The purpose of this Appendix is to demonstrate the utility of estimation-theoretic concepts for determining the information content in the visual cues provided by the extra-cockpit scene. The basic concepts of estimation and information are introduced first in generic terms. This is followed by an example which ranks the specific visual cues, defined by Eq. (1)-(3), for the visual approach to a hover tank.

The visual cues described by Eq. (1)-(3) are nonlinear functions of the aircraft state with respect to the helipad. Formally, these cues may be expressed as a nonlinear mapping

$$\underline{z} = f\ (\underline{x}) \qquad \text{(A.1)}$$

where $\underline{x} = (x_1, x_2, \ldots x_n)^T$ is the n dimensional aircraft state vector,

$\underline{z} = (z_1, z_2, \ldots z_p)^T$ is the p-dimensional visual cue vector,

and

$f = (f_1, f_2, \ldots f_p)^T$ are p nonlinear functions that map the, aircraft state vector $\underline{x}$ into the p individual visual cues z_i

However, the human pilot is limited in his ability to detect small changes in the visual cues z_i. Therefore, a visual threshold must be introduced in defining the perceived visual cues. Typically, two types of visual thresholds are considered - (1) a resolution or detection threshold, and (2) a discrimination threshold. A resolution threshold refers to the human ability to detect small changes in a variable from some explicit reference. The discrimination threshold describes the ability of the human to distinguish a small change in a variable in the absence of an explicit reference condition. The visual approach to a hover task must be performed in the absence of any explicit reference or desired condition. Therefore, discrimination threshold dominates the resolution threshold and must be considered in analyzing the perceptual process. The discrimination threshold in perceiving a visual cue z_i may be determined according to the Weber-Fechner law [30] as

$$\delta z_{iT} = \alpha_D \cdot z_i \tag{A.2}$$

where
$$\alpha_D = \frac{1}{30} \tag{A.3}$$

Furthermore, a threshold can equivalently be treated as an additive observation or perceptual noise v_{z_i} with a variance

$$\overline{v_{z_i}^2} \triangleq R_i = \left(\frac{\delta z_{i_T}}{2}\right)^2 \tag{A.4}$$

Equation (A.3) is based upon the results of an experiment conducted by Wewerinke [31-App.A] where v_{z_i} is treated as an observation error (or noise in perceiving z_i. Thus,

$$\underline{z}_p = \underline{z} + \underline{v}_z \tag{A.5}$$

where $\underline{z}_p$ is the equivalent perceived visual cue vector and $\underline{v}_z$ is the equivalent observation noise vector with covariance matrix

$$E\,(\underline{v}_z\,\underline{v}_z^T) = R = \begin{bmatrix} R_1 & & 0 \\ & R_2 & \\ & & \ddots \\ 0 & & R_P \end{bmatrix} \tag{A.6}$$

Eqs. (A.5)-(A.6) define the perceived cues available to the human pilot. The pilot's information processing task is to determine the best possible estimate of the aircraft state $\hat{\underline{x}}$ from the noisy perceived cues $\underline{z}_p$. Let the pilot's apriori (before perceiving the cue) estimate of $\underline{x}$ be $\bar{\underline{x}}$ having an uncertainty or error defined by the covariance matrix

$$E\ [(x - \bar{x})\ (x - \bar{x})^T] = M \qquad \text{(A.7)}$$

Then, a reasonable estimate $\hat{\underline{x}}$ that may be obtained by the pilot is the weighted-least-squares estimate which minimizes a quadratic cost [32]

$$J = \frac{1}{2}\ [(\underline{x} - \bar{\underline{x}})^T\ M^{-1}\ (\underline{x} - \bar{\underline{x}}) + (\underline{z}_p - H\underline{x})^T\ R^{-1}\ (\underline{z}_p - H\underline{x})] \text{(A.8)}$$

This estimate (which is identical to the minimum variance or maximum likelihood estimate for gaussian assumptions) is given by the standard Kalman filter equation

$$\hat{\underline{x}} = \bar{\underline{x}} + P\ H^T\ R^{-1}\ (\underline{z}_p - f(\bar{\underline{x}})) \qquad \text{(A.9)}$$

where $$H = \frac{\partial f}{\partial \underline{x}}\Big|_{x=\bar{\underline{x}}} \qquad \text{(A.10)}$$

is the linearized measurement matrix at $x = \bar{\underline{x}}$, the apriori state estimate, and

$$P = E\ [(\hat{\underline{x}} - \underline{x})\ (\hat{\underline{x}} - \underline{x})^T] \qquad \text{(A.11)}$$

is the estimation error covariance or dispersion matrix. Furthermore, it can be shown that

$$P^{-1} = M^{-1} + H^T R^{-1} H \qquad \text{(A.12)}$$

Since $H^T R^{-1} H$ is at least positive semi-definite (i.e., > 0), Eq. (A.12) shows that the estimation error covariance P after using the perceived cues $\underline{z}_p$ is never larger than M, the estimation error covariance before using the cues. Thus, the use of the perceived cues, on the average, reduces the uncertainty (i.e., error covariance) in the knowledge of the true state $\underline{x}$. Equivalently, the effect of cue utilization is to increase the amount of information available to the pilot about the instantaneous aircraft state. This interpretation is based on recognizing that the inverse of the error covariance or uncertainty is information. Thus M^{-1} and P^{-1} may be considered as the apriori (i.e., before cue utilization) and aposteriori (after cue utilization) information matrices, respectively. Thus, in the information context, the term $(H^T R^{-1} H)$ in Eq. (A.12) reflects the additional information about the state provided by the perceived cue vector $\underline{z}_p$. If the apriori information about the vehicle state $\underline{x}$ is poor, then $M^{-1} = 0$ and the information about the state after cue processing is

$$I = D^{*^{-1}} = (H^T R^{-1} H) \quad . \qquad (A.13)$$

It should also be noted that the dispersion matrix D^* is the error covariance matrix after measurement (assuming no apriori information) of the visual cues. Furthermore, D^* can be given a geometric interpretation in terms of constant likelihood (or uncertainty) hyperellipsoids

$$(\underline{x} - \hat{\underline{x}})^T D^{*^{-1}} (\underline{x} - \hat{\underline{x}}) = \ell^2 \qquad (A.14)$$

where ℓ is a constant.

The probability or likelihood that the true aircraft state $\underline{x}$ lies within a hyperellipsoid (A.14) depends upon the specific values for n and ℓ, and can be calculated. Thus, for a two state vector (n=2), the probability of finding the true $\underline{x}$ inside the ℓ=1 ellipse is 0.394, inside ℓ=2 ellipse is 0.865, and inside ℓ=3 ellipse is 0.989. These ellipsoids for ℓ=1,2,3 are usually referred to as 39, 86 and 99 percent likelihood ellipsoids.

The size and shape of the likelihood ellipsoid reflects the amount of information about the aircraft state contained in the perceived cue vector. Therefore, measures based upon the information matrix I or equivalently, the dispersion matrix D^*, may be used for quantifying and ranking the amount of information contained in a given set of visual cues. Typically, the following scalar measures are used:

$$J_1 = \text{Det } D^* = \prod_{i=1}^{n} \lambda_1 \qquad (A.15)$$

$$J_2 = \text{Tr } D^* = \sum_{i=1}^{n} \lambda_1 \qquad (A.16)$$

$$J_3 = \text{Max } \{\lambda_i(D^*)\} \qquad (A.17)$$

$$i = 1,2, \ldots , n$$

where λ_i, i = 1,2, .. , n are the eigenvalues of D^*. Note that the square roots of the eigenvalues $\sqrt{\lambda_i}$ correspond to the magnitudes of the principal axes of the hyperellipsoid along their respective eigenvector directions. The determinant of D^*, (J_1), is a measure of volume of the one sigma(ℓ=1) likelihood ellipsoid. Trace D^*, (J_2), gives a measure of the average mean-squared estimation error, and the maximum eigenvalue of D^*, (J_3), represents the largest principal semi-axis of the hyperellipsoid and hence the worst case estimation error variance.

The ranking of the visual cues based upon the dispersion matrix implicitly takes into consideration the effects of cue sensitivities and thresholds. This is apparent from the definition of the dispersion matrix as

$$D^* \triangleq H^T R^{-1} H$$

where $H = \frac{\partial f}{\partial \underline{x}}$ is the sensitivity matrix and R is a diagonal observation noise matrix with diagonal elements reflecting the discrimination threshold levels for each of the available cues.

A.1 Information Ranking of the Visual Cues - An Example.

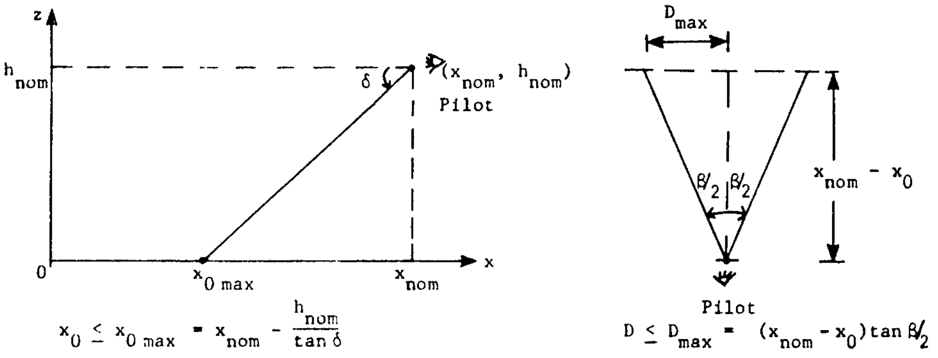
An analytical framework for determining and ranking information content in the visual cues has been presented. Application of this methodology for evaluating the visual cues during the approach to a hover task is discussed next. The visual cues available to the pilot are limited to the geometric cues described by Eq. (1)-(3), which are repeated here for convenience

$$\ell_{x_0} = \frac{c\ h}{x - x_0} \tag{1}$$

$$w_{x_0} = \frac{c\ D}{x - x_0} \tag{2}$$

$$\eta = \tan^{-1} \frac{h}{D} \tag{3}$$

If the rectangular helipad shown in Figure 2 is the only source of visual cues to the pilot, then a finite set of cues corresponding to the perceived lengths, widths and inclinations may be used by the pilot for the state estimation task. However, for most realistic situations, the pilot has access to far more visual cues than those provided by a helipad alone. Thus, useful cues may be available from a much larger field-of-view than that covered by the helipad image. Conceptually, a larger field-of-view in azimuth and elevation for the visual scene may be represented by superimposing a grid structure on the ground plane with lines parallel to the longitudinal and lateral edges of the helipad. Mathematically, this is equivalent to allowing a range of discrete values for x_0 and D in Eq. (1)-(3) for the visual cues. Figure A.1 shows the elevation (side view) and plan (top-down) views of the helipad approach scenario. The pilot's view outside the window is limited in elevation (δ) and azimuth (β) by virtue of the canopy configuration.



Side View

Top View

Figure A.1. Field-of-View Limitations

For given δ and β, the values of x_0 and D must be bounded:

$$x_0 < x - \frac{h}{\tan\delta} \tag{A.18}$$

and
$$D < 2\ (x - x_0)\ \tan\frac{\beta}{2}$$

where (x,h) are the longitudinal range and altitude of the helicopter with respect to the helipad center.

For a given x_0 and D, Eq. (1)-(3) represent three visual cues that may be used by the pilot for estimating his ground range x and altitude h. At least two cues are needed to estimate x and h. Thus, the pilot may select any of the following four sets of cues:

$$C_0 \triangleq (w_{x_0}, \eta) \tag{A.19a}$$

$$C_1 \triangleq (\ell_{x_0}, w_{x_0}) \tag{A.19b}$$

$$C_2 \triangleq (\ell_{x_0}, \eta) \tag{A.19c}$$

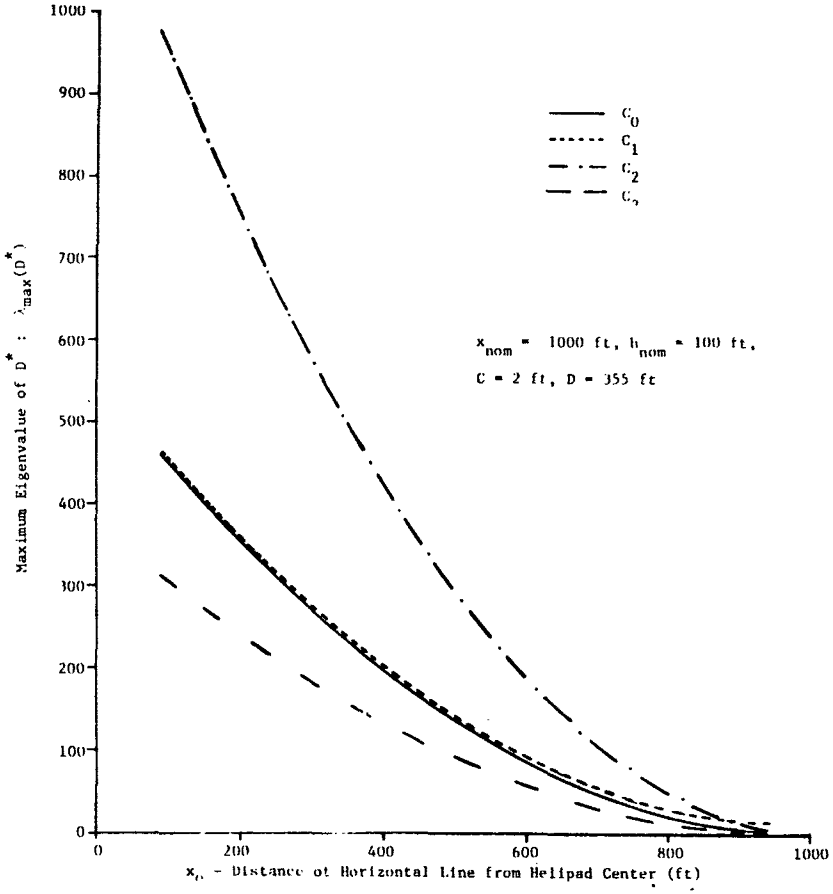


Figure A-2. λ_{max} (D^*) versus x_0

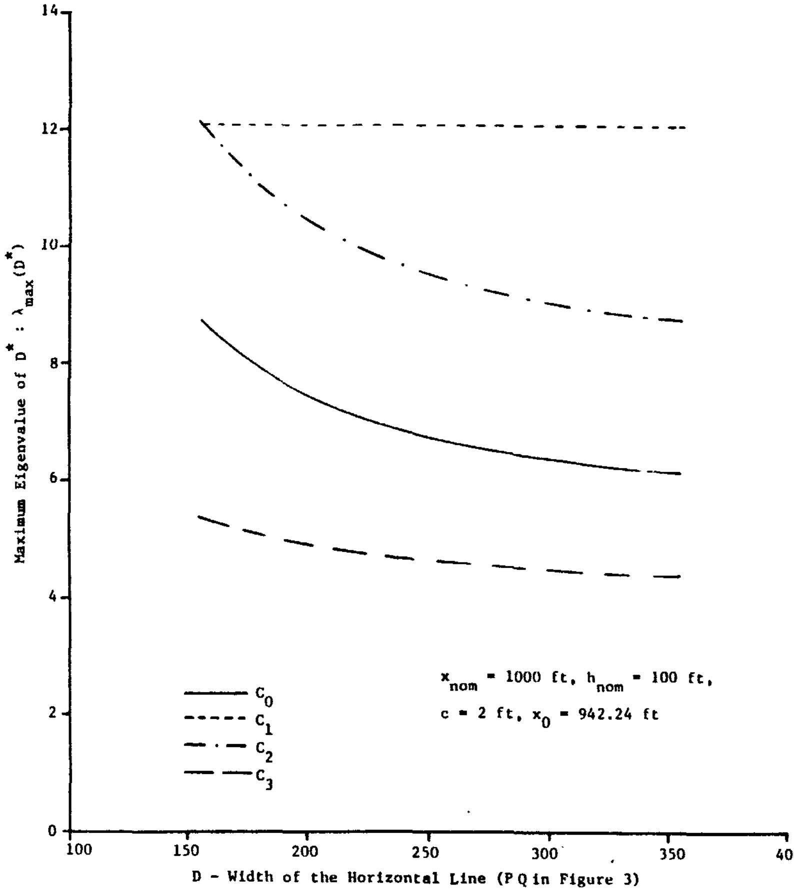


Figure A-3. λ_{max} (D^*) versus D for a finite or continuum of values of i, that may correspond to horizontal grid line elements of varying length D_i. The use of additional extra-cockpit visual cues provides more information to estimate the helicopter state and should result in reduced estimation error covariance and thus improved state estimates. Just how many visual cues are sufficient would depend upon the nature of the curve describing the trade-off between the number of cues (abscissae) and the cumulative information (ordinate). Usually, the number of cues selected is truncated on the knee of the cumulative information versus the number of cues curve where the amount of incremental information provided by adding a cue is not justified by the cost or effort of acquiring and processing that cue.

REFERENCES

1. Wierwille, W. W., and Williges, B. H., "Naval Air Test Center Technical Report," Systemetrics, Inc., Report No. SY-27R-80, March 1980.

2. Cooper, G. E., and Harper, R. P., Jr., "The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities," NASA TN D-5153, April 1969.

3. McRuer, D. T., and Jex, H. R., "A Review of Quasi-Linear Pilot Models," IEEE Trans., Vol. HFE-8, No. 3, September 1967, pp. 231-249.

4. Kleinman, D. L., Baron, S. and Levison, W. H., "A Control Theoretic Approach to Manned-Vehicle Systems Analysis," IEEE Trans. on Auto Control, AC-16, 1971, pp. 824-832.

5. Phatak, A. V., and Karmali, M. S., "Pilot Model-Based Analysis of Typical Helicopter-Mission Segments," AMA Report #81-40, November 1981.

6. Forgus, R. H., and Melamed, L. E., "Perception - A Cognititive-State Approach," McGraw-Hill, Second Edition, 1976.

7. Haber, R. N. and Hershenson, M., "The Psychology of Visual Perception," Holt, Rinehart and Winston, 1973.

8. Carterette, E. C., and Friedman, M. P., "Handbook of Perception," Vol. IX Perceptual Processing, Academic Press, NY 1978.

9. Hochberg, J. E., "Perception," second edition, Prentice-Hall, Inc., NJ 1978.

10. Matheny, W. G., et al, "An Investigation of Visual, Aural, Motion and Control Movement Cues," Naval Training Device Center, Technical Reort NAVTRADEVCEN 69-C-0304-1, April 1971.

11. Graham, C. H., "Vision and Visual Perception," New York, John Wiley and Sons, 1966.

12. Gibson, J. J., The Perception of the Visual World, Riverside Press, Cambridge, MA, 1950.

13. Moen, G. C., DiCarlo, D. J. and Yenni, K. R., "A Parametric Analysis of Visual Approaches for Helicopters," NASA TN D-8275, December 1976.

14. Havron, M. D., "Information Available from Natural Cues During Final Approach and Landing," Human Sciences Research, Inc., Report HSR-RR-623-MK-X, March 1962.

15. Calvert, E. S., "A Quantitative Analysis of the Ground Indications Seen by the Pilot When Using Visual Aids for Landing in Good and Bad Visibility," Royal Aircraft Establishment, Farnborough, Hants, Report EL 1459, May 1949.

16. Calvert, E. S., "Visual Aids and Their Effect on Landing Success and Safety," Royal Aircraft Establishment, Farnborough, Report No. EL 1486, October 1955.

17. Baron, S., Lancraft, R., and Zacharias, G., "Pilot/Vehicle Model Analysis of Visual and Motion Cue Requirements in Flight Simulation," Bolt Beranek and Newman, Inc., Report No. 4300, January 1980.

18. Naish, J. M., "Control Information in Visual Flight," Seventh Annual Conference on Manual Control, NASA SP-281, June 1971.

19. Sinacori, J. B., "V/STOL Ground-Based Simulation Techniques," USAAVLABS Technical Report 67-55, November 1967.

20. Baron, S. and Berliner, J., Manmod 1975; Human Internal Models and Scene-Perception Models, U.S. Army Missile Command Technical Report RD-CR-76-3, September 1975.

21. Grunwald, A. and Merhav, S. J., "An Analytical Model for Visual Field Control - A Parametric Study," Israel Institute of Technology, Report TAE-217, June 1974.

22. Wewerinke, P. H., "A Theoretical and Experimental Analysis of the Outside World Perception Process, Fourteenth Annual Conference on Manual Control, NASA SP-2060, November 1978.

23. Lane, J. C. and Cumming, R. W., "Pilot Opinions and Practices on the Approach to Landing: A Questionary Survey Among Australian Civil and Military Pilots," Department of Supply, Australian Defence Scientific Service, Aeronautical Research Laboratories, Report ARL/H.E. 1, April 1959.

24. Richalet, J., Rault, A., Testud, J. L., and Papon, J., "Model Predictive Heuristic Control: Application to Industrial Process Control, Automatica, Vol. 14, 1976, pp. 413-419.

25. Richalet, J., "General Principles of Scenario Predictive Control Techniques," Proceedings, 1980 Joint Automatic Control Conference, San Francisco, CA, August 1980, FA9-A.

26. Mehra, R. K., Kessel, W. C., Rault, A., Richalet, J., and Papon, J., "Model Algorithmic Control Using IDCOM for the F-100 Jet Engine Multivariable Control Design Problem," in Alternatives for Linear Multivariable Control with Turbofan Engine Theme Problem, Edited by Sain, Peczkowki, and Melsa, National Engineering Consortium, Inc., Chicago, 1978, pp. 317-350.

27. Reid, J. G., Chaffin, D. E. and Silverthorn, J. T., "Output Predictive Algorithmic Control: Precision Tracking with Application to Terrain Following," J. Guidance & Control, Vol. 4, No. 5, pp 502-509, AIAA Paper #81-4242.

28. Teper, G. L., "An Assessment of the "Paper Pilot" - An Analytical Approach to the Specification and Evaluation of Flying Quantities," AFFDL TR-71-174, June 1972.

29. Levison, W. H., Elkind, J. I., and Ward, J. L., "Studies of Multivariable Manual Control Systems, A Model for Task Interference," NASA Cr-1746, May 1971.

30. Luce, R. D. and Galanter, E., "Discrimination, in Handbook of Mathematical Psychology," Vol. I, John Wiley & Sons, New York 1963.

31. Wewerinke, P. H., "Visual Scene Perception Process Involved in the Manual Approach," National Aerospace Laboratory, The Netherlands, Report #NLR TR 78130U.

32. Bryson, A. E. Jr. and Ho, Y. C., "Applied Optimal Control," Blaisdell Publishing Co. 1969.

PURDUE UNIVERSITY
SCHOOL OF AERONAUTICS AND ASTRONAUTICS

Pilot Model Determination Using Parameter Identification Methods on Hovering VTOL Flight Data*

by

Dominick Andrisani II
Simon M. Bourne
Ching-Fu Gau

of

Purdue University
School of Aeronautics and Astronautics
West Lafayette, IN 47907

A paper presented at the "Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics".
Edwards AFB, CA, 19-21 January 1982.

*This paper is based on research being funded by the Naval Air Development Center, Warminster, PA under contract #N622-69-81-R-0729 with Mr. John Clark, Jr. as contract monitor.

1. Introduction. This paper presents highly intriguing preliminary results concerning the mathematical model of human pilots hovering over a moving landing pad in an X-22A VTOL aircraft. The results are in the form of a time domain mathematical model which was determined from transient responses of the X-22A. The pilot, using both a head-up-display for landing pad position information and the outside visual scene for attitude information, was asked to precisely hover over a simulated landing pad which would move in semi-random step-like jumps. Since the X-22A (with its translation rate control system) was also driven by simulated ship wake turbulence and natural turbulence, the pilot's task was to track step inputs while correcting for random disturbances.

Unlike many other research efforts the research reported here is based upon time domain parameter identification methods. The advantage of the time domain approach is that it effectively duplicates the situation observed in flight, i.e. the need to make discrete maneuvers while minimizing disturbances. The assumption of stationarity is never made and for good reason since very few piloting tasks are even remotely stationary or time invariant in nature.

The proposed pilot model encompasses two primary control policies which were observed in the flight data. The first is a bang-bang control policy for amplitude limited gross maneuvers while the second policy is an error minimization/disturbance rejection linear control policy. Efforts are under way to identify the unknown parameters of the model from the X-22A flight data using state and parameter estimation techniques. Identified pilot models for various X-22A dynamic configurations will then be correlated with known vehicle handling qualities in order to formulate a control system design procedure directly incorporating handling qualities considerations.

2. Flight Task Description. Evolving Naval tactics are today forcing naval helicopters airborne in all weather and on ever smaller ships. The flight data upon which this research is based was generated in an attempt to study the problem of visual landing of a helicopter or VTOL on a small non-aviation ship in bad weather. Funded through the Naval Air Development Center, the Calspan Advanced Technology Center devised a flight experiment (References 1, 2) using the X-22A variable stability aircraft, Figure 1, in which the pilot was instructed to remain positioned over a moving landing pad in the presence of artifically generated ship wake turbulence and some unavoidable natural atmospheric turbulence. Since the task was a visual one, the pilot could observe absolute position, velocity, and altitude by viewing the outside world through the cockpit canopy. The position of the landing pad was presented through a head-up-display (HUD) using a display format shown in Figure 2. The information presented was symbolic since the field of view of the HUD effectively precluded a pictorial presentation of the landing pad and the ship superstructure. Salient features of the display are the round fixed aircraft symbol (scaled to a 10 foot diameter) and the altitude ladder with rung separation scaled to 10 feet. The longitudinal and lateral displacement from the landing pad (square symbol) are presented in plan view in a heading axis system. Referring to Figure 2, closure with the pad would require forward and right stick. Height above the landing pad is depicted by the separation of the double dumbell and the aircraft symbol, in effect an elevation view of the vertical situation. As in the x-y situation presentation, the control is in the fly to sense, that is, the dumbell symbol is the landing pad.

The pursuit tracking task presented to the pilot via the HUD involved discrete semi-random step-like jumps in landing pad position. In the first portion of any given pilot evaluation the pad position changed once every twenty or thirty seconds in either x position separately, y separately or both x and y positions simultaneously. Since the vertical axis flight control system employed altitude hold the pilot's task of moving to the new landing pad and then hovering accurately over the pad in the presence of disturbances was overall a two axis task with definable portions of one axis activity. In the second portion of any pilot evaluation the task involved the same sequence of step-like landing pad movements applied in the Y and Z pad positions. This portion of the pilot's evaluation was therefore a three axis task since X stationkeeping remained a demanding chore.

After a step change in landing pad position the pilot performed a gross maneuver, examples of which are discussed later, in which the pilots objective was to rapidly reposition over the new pad position. Once repositioning over the pad was carried out the pilot attempted to hover over the pad to within an accuracy of one-half symbol width (± 5 feet).

3. Flight Control System. The test vehicle, the X-22A operated by the Calspan Advanced Technology Center under sponsorship from the Naval Air Development Center, was configured so that the response to a steady pitch or roll stick input generated a steady state inertial translational velocity in the x or y direction respectively. This translational rate command (TRC) system, also called a velocity command control system, was mechanized in both, the pitch and roll axis as shown in Figure 3. For

a step control stick command the response of vehicle translational velocity and attitude is shown in Figure 4. Note that the X-22A must change attitude in order to tilt the thrust vector in order to translate either side-to-side or fore and aft.

The TRC system was mechanized using feedback of inertial velocity ($K_{\dot{x}}$, $K_{\dot{y}}$), attitude (K_θ, K_ϕ) and angular rates (K_q, K_p). The resulting closed loop transfer functions for translational velocity ($\dot{x}$ or $\dot{y}$) and attitude (θ or ϕ) are given below.

$$\frac{\dot{x}}{\delta_{ES}} = \frac{-K_e M_{\delta_e} g}{(s-X_u)(s^2+(K_q M_{\delta_e}-M_q)s+K_\theta M_{\delta_e}) + g(M_u+K_x \cdot M_{\delta_e})}$$

$$= \frac{-K_{\dot{x}_c}}{(1+s/\lambda)(1+(2\zeta/\omega_n)s+(s/\omega_n)^2)} \qquad (1)$$

$$\frac{\theta}{\delta_{ES}} = \frac{s-x_u}{g} \; \frac{K_{\dot{x}_c}}{(1+s/\lambda)(1+(2\zeta/\omega_n)s+(s/\omega_n)^2)}$$

The system gains were selected so that the dynamic response and stick force system characteristics were identical in both pitch and roll axes.

The yaw axis flight control system was not a significant contributor to the flight experiment and the pilot was able to keep his feet off the rudder petals. While the altitude hold vertical axis flight control system was of significance in the Calspan flight test program it is not of consequence for this paper since we deal here with only X and Y maneuvering at constant altitude.

4. The Data Base. The primary purpose of the Calspan/NADC flight experiment was to determine the effect of variations in command path gain (steady state translational velocity per inch of stick deflection, $K_{\dot{x}}$), and path mode time constant (λ in Equations 1) upon vehicle handling qualtities. In order to explore these factors the attitude and angular rate feedback gains were adjusted on the majority of the flight configurations so as to maintain an inner loop attitude system with natural frequency of 2.5 radians per second and damping ratio of .8. The velocity feedback gains and forward loop gains were varied to realize path mode time constants ranging from 1.5 to 4 seconds and command path gains from 3 to 12 ft/sec/in. The inherent flexibility of the X-22A's variable stability control system made this task possible through a bank of pilot adjustable potentiometers located in the cockpit.

Approximately 86 piloted evaluations of the type described in the last section have been obtained by the authors from the Naval Air Development Center on fourteen magnetic tapes. Each tape contains time histories of all signals pertinent to the experiment.

Following each evaluation the pilots (there were two pilots in the available data base) were instructed to provide Cooper-Harper pilot ratings and to provide detailed commentary. This data is available in Reference 2.

5. Research Objectives. The objectives of this study are listed below:

i To formulate a useful mathematical model of the human pilot while performing a discrete tracking task and station keeping duties in a hovering VTOL (X-22A).

ii To employ time domain parameter identification methods in order to determine the numerical values of all unknown terms in the pilot model. This step is to be repeated for a broad range X-22A dynamics in the class of translational rate command systems.

iii To determine useful measures of closed loop pilot/vehicle performance.

iv to correlate identified pilot models and closed loop vehicle performance measures with known handling qualities.

v To generalize the analysis so as to develop a control system design tool capable of predicting closed loop pilot/vehicle handling qualities.

6. Research Results. The results to date include the formulation of a time domain nonlinear mathematical model of the human pilot based upon a detailed and continuing analysis of the X-22A flight data. This model is presented and justified in the remainder of this section. Objectives ii- v remain to be satisfied in this continuing research program.

6.1 Typical Flight Data. Figures 5 through 8 contain typical flight data showing the pilot/vehicle response to a step change in x-longitudinal landing pad position. Each figure contains three time history plots. The top one labeled X ERROR shows the error between the x-pad position and the X-22A position. The steplike jump in X ERROR is the result of the sudden landing pad position change described in Section 2. The center plot, labeled DE, is the fore-and-aft position of pilot's stick, i.e. the pilots primary output in this x-positioning task. The lower plot, labeled Q, is the resulting pitch rate of the X-22A. Table 1 shows the path mode time constant λ, the command path gain $K_{\dot{x}}$, and the Cooper-Harper pilot rating for the data presented in the Figures.

Table 1

Figures	Configuration	λ(sec)	$K_{\dot{x}}$(ft/sec/in)	Cooper-Harper PR
5, 6	208A	2.75	5.5	6
7, 8	206B	2.0	6.0	4

Careful examination of these time histories show that following the step landing pad change the pilot moves the pitch controller with two types of policies. The first stick motion policy involves bang-bang type of motion in which the pitch control is moved to a pilot perceived control limit, held there for typically 5 seconds, moved to a limit in the opposite direction, held there for another 5 seconds and then returned to zero. This strategy is certainly based upon the pilots desire to move as rapidly as possible to the new landing pad position and stop over the pad without overshoot. The fact that the control used is of limited magnitude should be of no surprise if it is recalled that in a tilt-to-translate vehicle such as the X-22A there are large pitch attitude excursion associated with translation and it seems reasonable that the pilot will tolerate these only up to some maximum values.

The second policy apparent in the pilot's stick consists of smaller amplitude oscillations at frequencies less than 1 Hertz. Examination of the pitch rate, Q, plot shows strong correlation between the high frequency stick and pitch rate activity. This type of pitch stick activity is probably due to a combination of pilots inability to accurately carry out his desired control policy (i.e. motor noise) and the pilots desire to surpress the X-22A's response to natural turbulence and synthetically generated shipwake turbulence.

In summary two types of pilot strategies can be hypothesized from Figures 5-8. The first is a gross acquisition maneuvering characterized by large amplitude bang-bang control motion. The second is a lower amplitude higher frequency disturbance rejection. The pilot model proposed subsequently has been designed to deal with both strategies.

6.2 Bang-Bang Optimal Control. For the germane problem of minimizing the time to reposition over the landing pad subject to amplitude limited control the theory of optimal control leads to a rather elegant bang-bang optimal control policy. The previous section and Figures 5-8 strongly suggest exactly this policy was employed by the X-22A pilot.

If the second order portion of equations which characterize primarily the attitude response characteristics is ignored the resulting differential equations describing x-position and translational velocity are of second order. With this reduced order model the optimal control can be computed to track a step change in pad position in minimum time with control limited to plus or minus one inch. The optimal bang-bang control computed in this way is shown in Figure 9 which is most directly comparable to the in-flight measured situation shown in Figure 6. Comparison of these two figures shows reasonably accurate agreement on switching times and gives strong evidence towards establishing the veracity of the optimal bang-bang pilot control policy postulated earlier.

The next sub section proposes a model structure which integrates the two pilot policies which have just been postulated.

6.3 Proposed Pilot Model. A pilot model is described in this section which includes the two strategies hypothesized in previous sections. The model shown in Figures 10-12 is based upon the following assumptions.

1. The pilot, upon seeing a step change in pad position, $x_{ref,}$ will compute a bang-bang control policy δ_c, so as to reposition over the pad in minimum time. The maximum allowable positive and negative limits on control activity (δ_1, δ_2 of Figure 12) are not constrained to be the same allowing for a possible pilot preference to translate faster forward with nose down and the ground in sight.
2. Once the bang-bang control policy is formulated the pilot determines mentally the desired aircraft motion i.e. the motion the aircraft would follow as a result of his bang-bang control policy. This becomes the desired aircraft trajectory ($X_D(t)$).
3. Because the pilot is unable to perfectly carry out his desired control policy there will be errors, motor noise m(t), in the actual positioning of his control.
4. The pilot's output is delayed because of perceptual and neuromuscular pure time delays and his actual movement is assumed to be modeled as a first order lag. The latter seems necessary because of the obviously finite slopes of the pilots output, DE, at the switching times apparent in Figures 5-8.

5. The actual X-22 receives inputs not only from the pilot but from at least two sources of disturbances, natural atmospheric turbulence and synthetic ship wake turbulence. In the flight experiment the latter disturbance was deliberate and measured while the former is random, unmeasured, and therefore unknown.

6. The pilot, sensing the disturbances, will tend to generate a control which minimizes the differences between the actual X-22A responses, $X_A(t)$, and the desired X-22A responses, $X_D(t)$ by feeding back the response error through a feedback gain matrix, K_p.

6.4 Estimation of Pilot Models. Given the pilot model discussed above and described in Figures 10-12 it is possible to determine the unknown parameters of the model by using the time history measurements of actual pilot/vehicle performance and modern state and parameter estimation techniques. Efforts by the authors in this direction are continuing.

The unknown parameters which must be determined from the flight data are listed on Figures 11 and 12. It should be noted that the switching times and magnitudes in the pilots bang-bang control policy are estimated from the flight data and are not, in the general case, the result of forcing optimality. Therefore, the pilot's desired trajectory model may not be optimal in any sense and in fact may not even guarantee exact positioning over the landing pad. If this is determined to be a problem additional pilot integral feedback paths may be required to insure proper vehicle positioning with respect to the actual landing pad.

The extension of this approach to the multi-axis task can be accomplished by identifying "urgency functions" or attention switching boundaries from the flight data. These "urgency functions" developed by Onstott and Faulkner in Reference 3 allow for pilot attention switching from one axis task, eg x-positioning, to another axis task, eg. y-positioning. Efforts in this direction are planned.

7. Conclusions. A pilot model has been proposed based upon a preliminary analysis of hovering X-22A flight data performing a discrete tracking task in the presence of disturbances. The pilot model encompasses both a bang-bang control policy for amplitude limited gross maneuvers and an error feedback model for disturbance rejection. Attempts are underway to use this model form to identify the unknown parameters of the model from the X-22A flight data using state and parameter estimation techniques. Extension to multi-axis tasks are possible by identification of urgency or attention functions from the flight data.

8. References

1. Radford, R.C. and Andrisani, D. II, "An Experimental Investigation of VTOL Flying Qualities Requirements in Shipboard Landings", presented at the AIAA 7th Atmospheric Flight Mechanics Conference, Danvers, Mass. August 11-13, 1980.
2. Radford, R.C., Andrisani, D. II, and Beilman, J.L., "An Experimental Investigation of VTOL Flying Qualities Requirements for Shipboard Landings", NADC-77318-60, August 1981.
3. Onstott, E.D. and Faulkner, W.H., "Prediction, Evaluation, and Specification of Closed Loop and Multi axis Flying Qualities", AFFDL-TR-78-3, February 1978.

GENERAL SPECIFICATIONS

DIMENSIONS

Length	39.57 ft	
Height	20.69 ft	
Tread	8.0 ft	
Wing	Front	Aft
Area	139 sq ft	286 sq ft
Span	22.97 ft	39.24 ft
Aspect Ratio	3.86	5.38

ENGINE RATINGS

SHP	SLS	Thrust	rpm	Min.
1250	Mil.	154	19,500	30
1050	Nor.	132	19,500	Cont.

POWER PLANT

No. & Model	(4) YT58-GE-8D
Mfr.	General Electric Co.
Type	Free Power Turbine
Reduction Gear Ratio	0.133
Prop Mfr.	Hamilton Standard
Prop. Dia.	84 in.
No. of Blades	3
Tail Pipe	Fixed Area

WEIGHTS

Loading	lb
Empty	11,622
Gross	15,287
Max Takeoff	18,420
Max Landing	15,287

FUEL

No. Tanks	Gal	Location
1	465	Fuselage

Fuel Grade JP-4 or JP-5

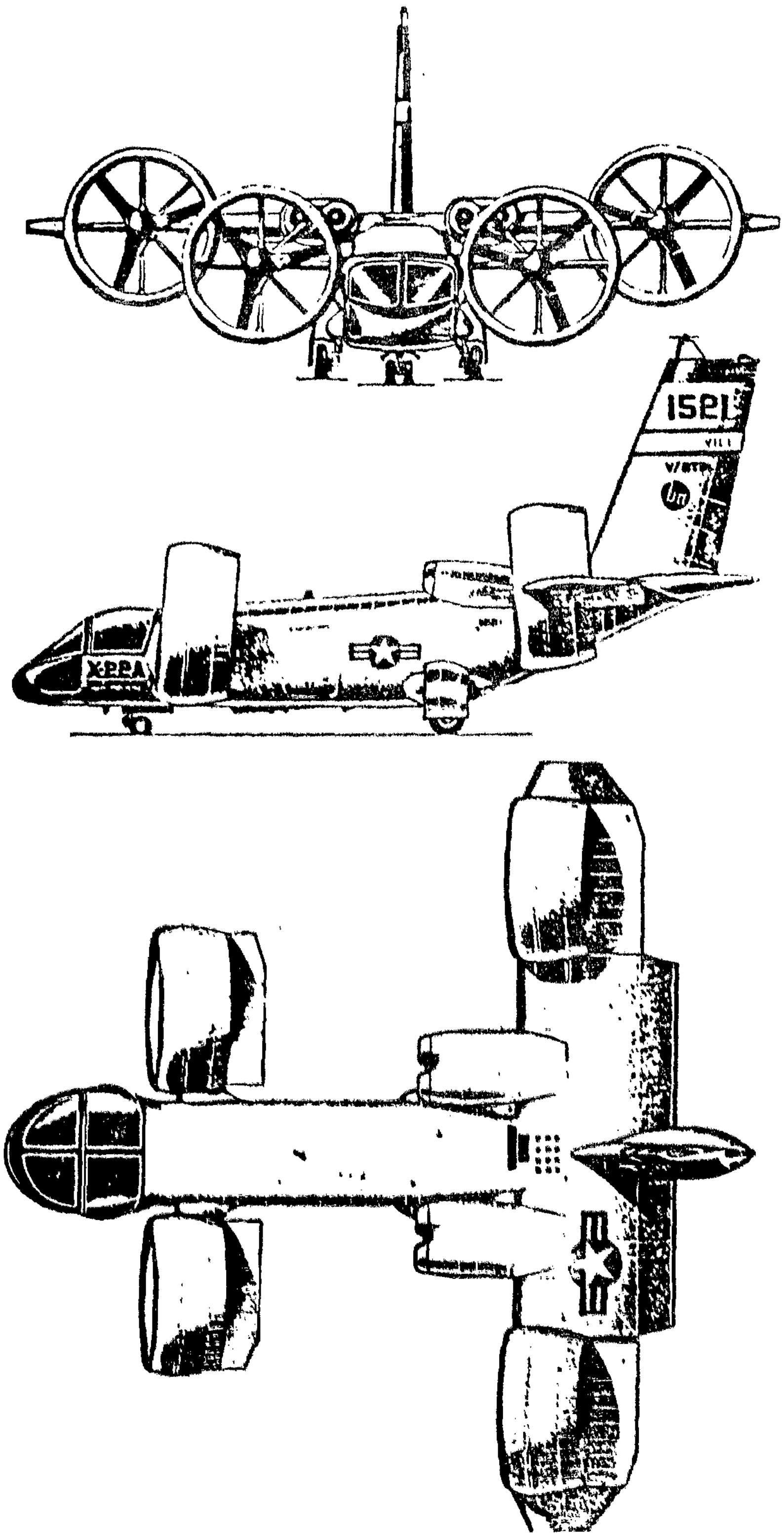


FIG. 1 X-22A AIRCRAFT, 3-VIEW

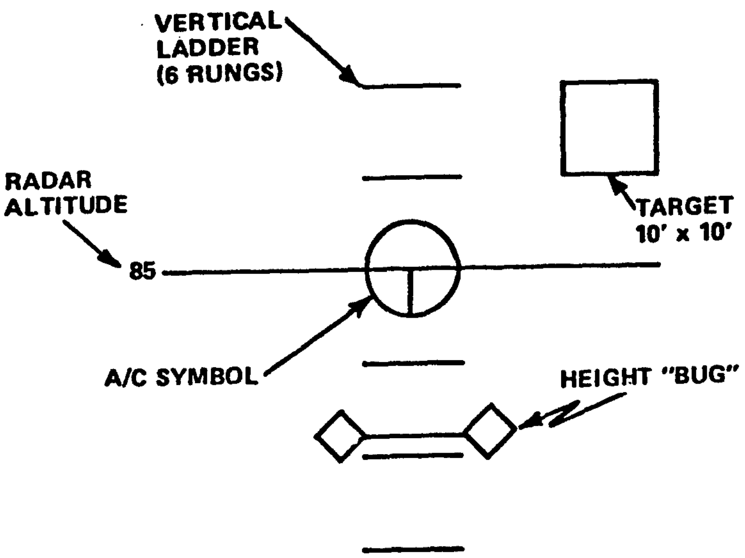


Fig. 2 HEAD-UP-DISPLAY SYMBOLOGY FOR HOVERING TASK

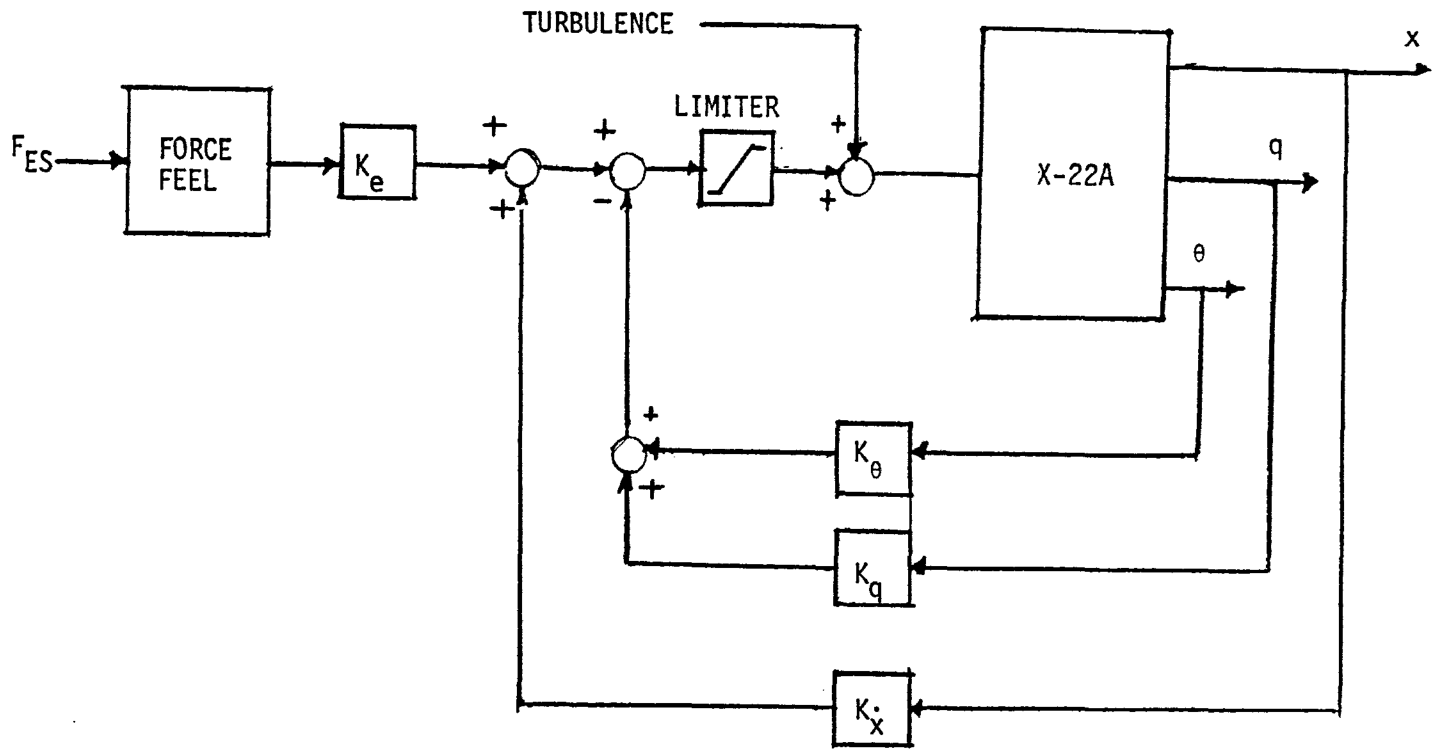


FIG. 3 LONGITUDINAL/LATERAL AXIS FLIGHT CONTROL BLOCK DIAGRAM

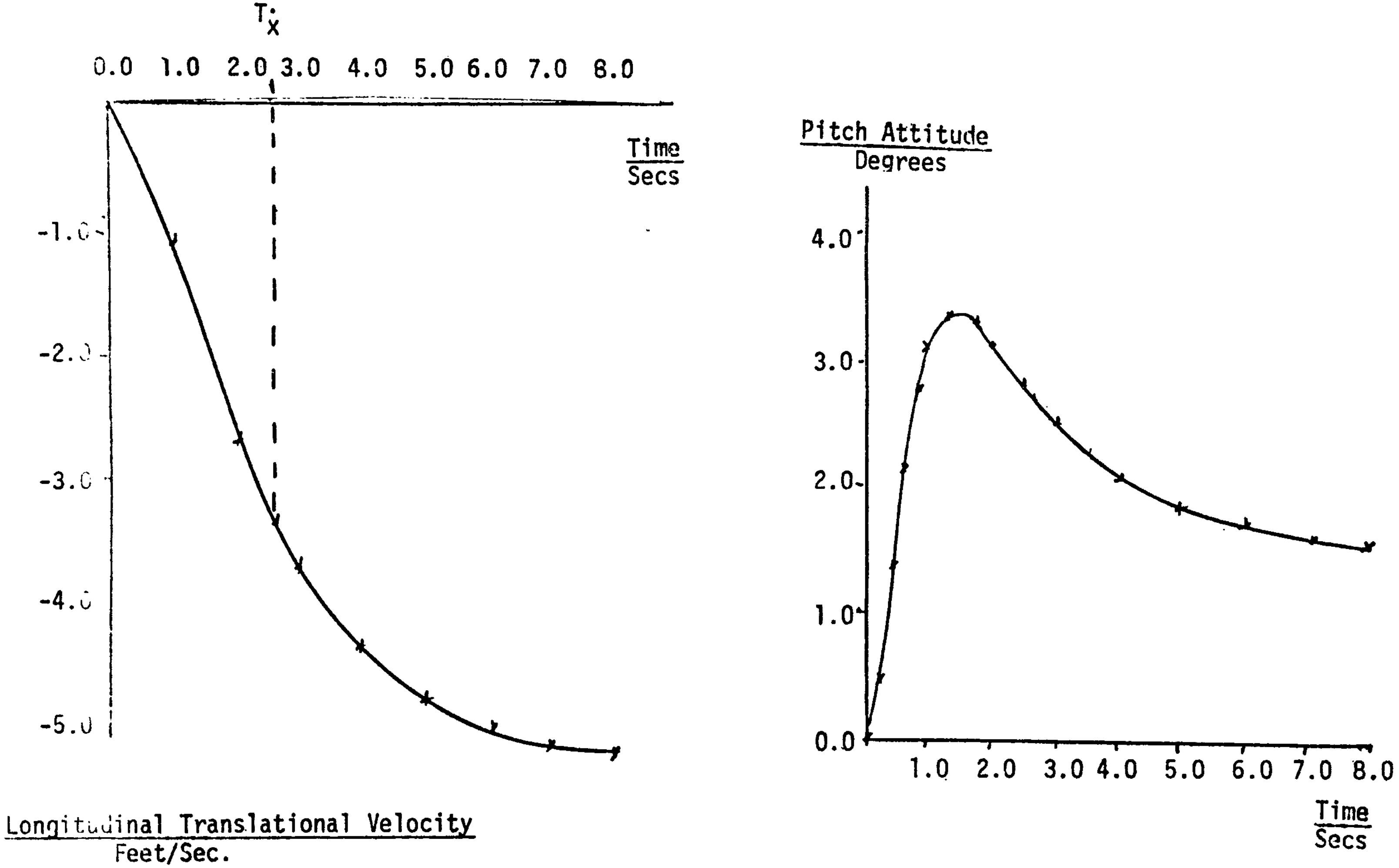


FIG. 4 RESPONSE TO ONE INCH PITCH STICK DEFLECTION (+VE DEFLECTION AFT) FOR X-22A FLIGHT CONFIGURATION 208A

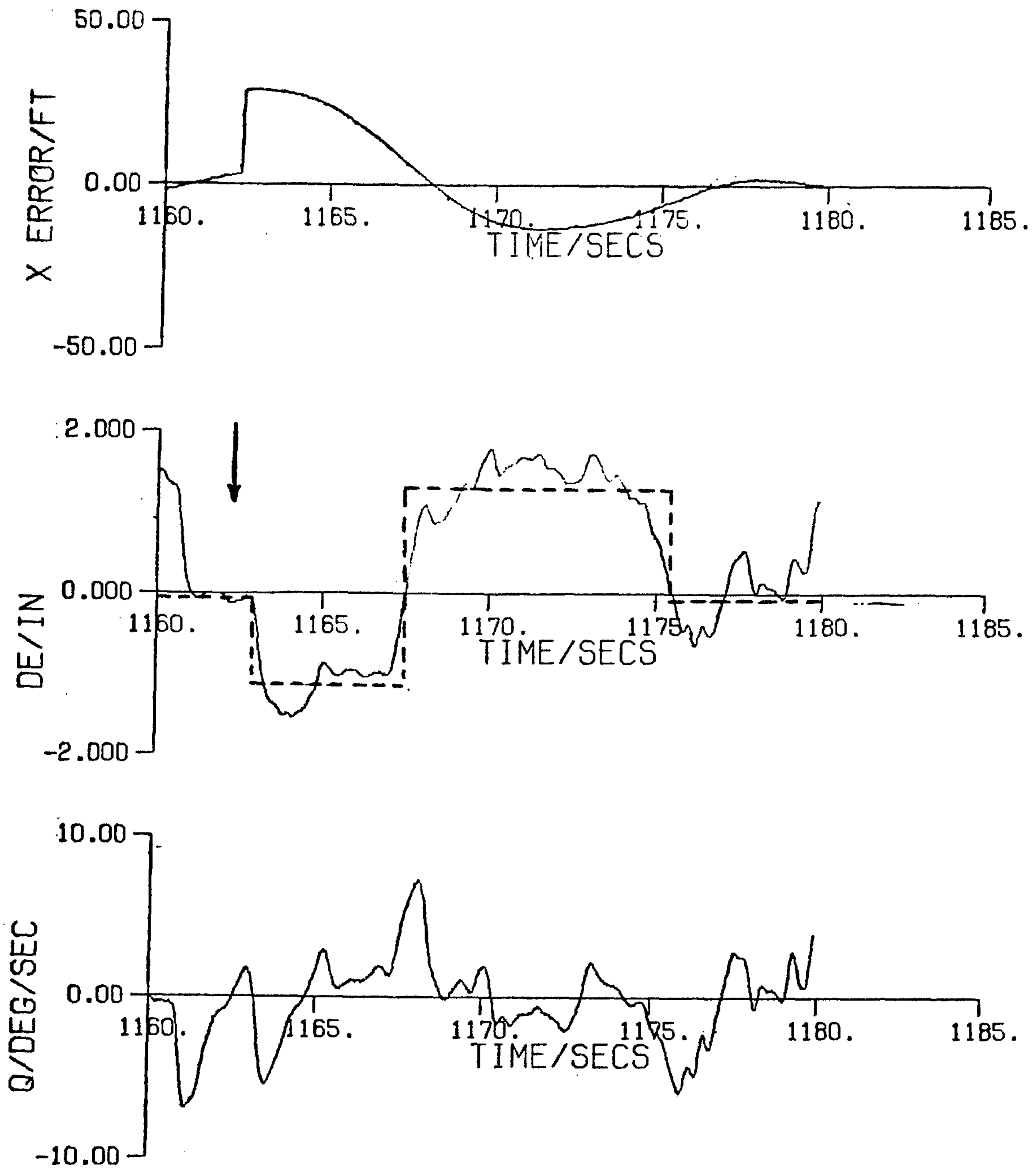


FIG. 5 X-22A FLIGHT DATA HISTORIES
ID:208A
F.C.: 817

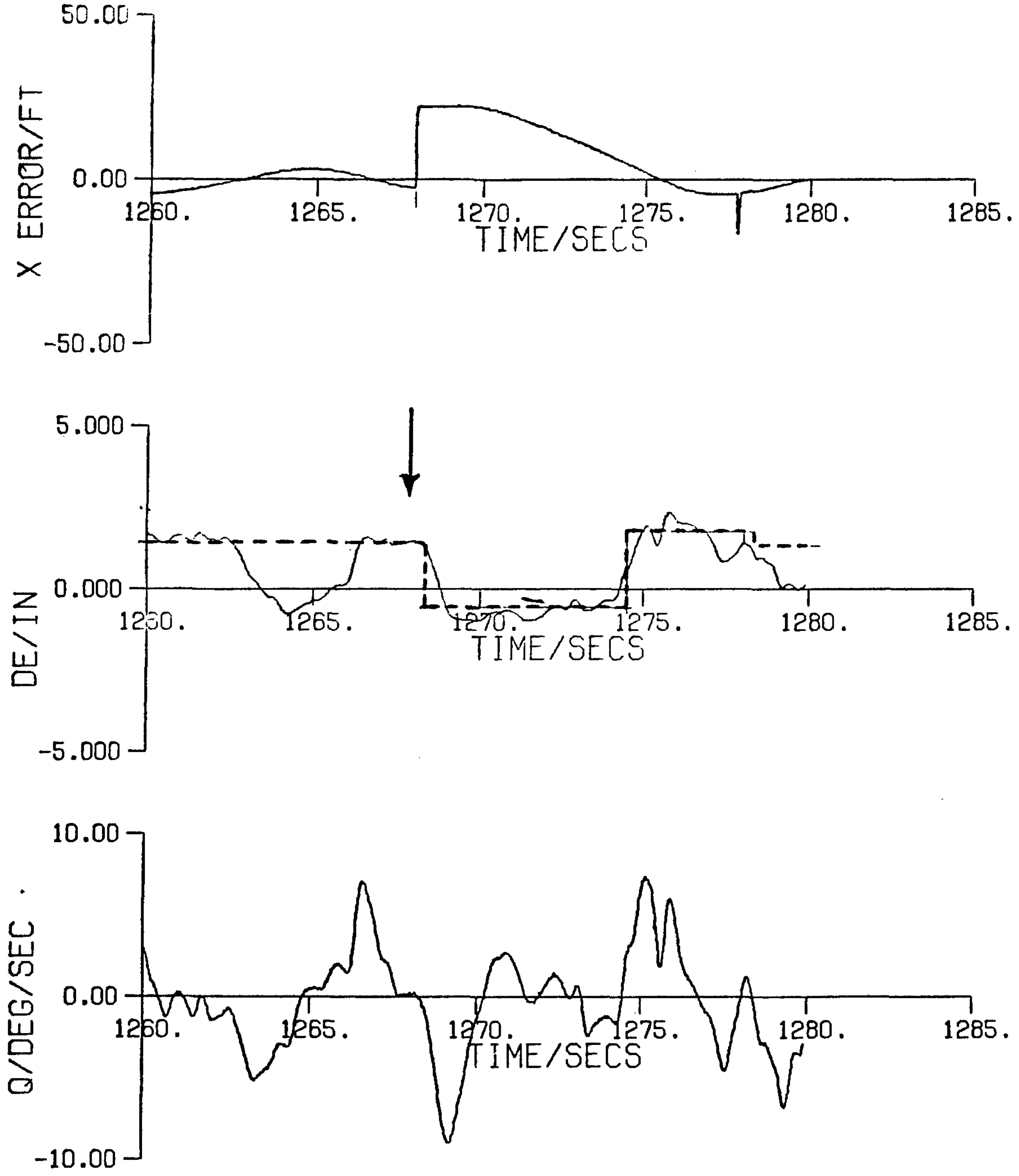


FIG. 6 X-22A FLIGHT DATA HISTORIES
ID:208A
F.C.:811

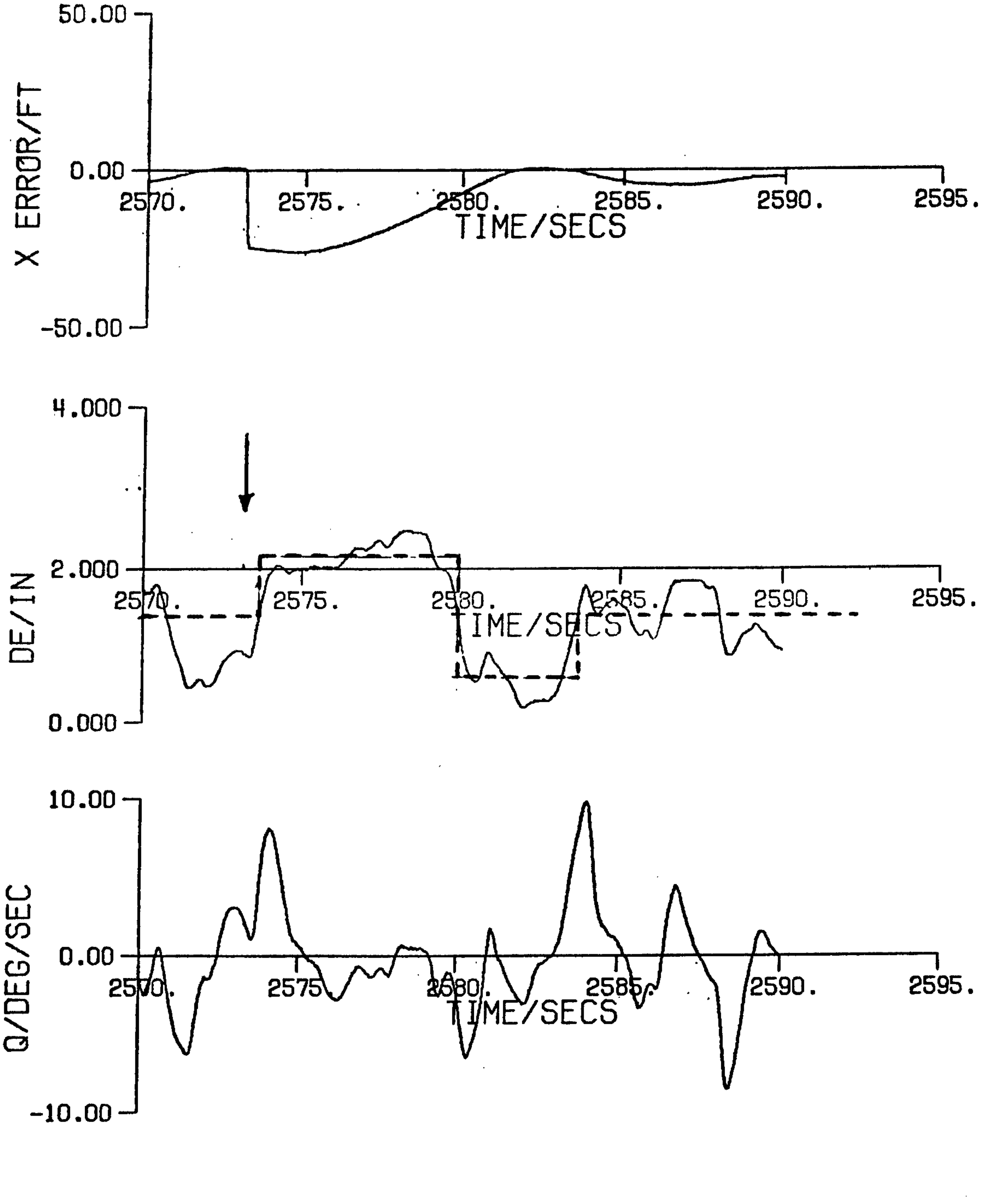


FIG. 7 X-22A FLIGHT DATA HISTORIES
ID:206B
F.C.:622

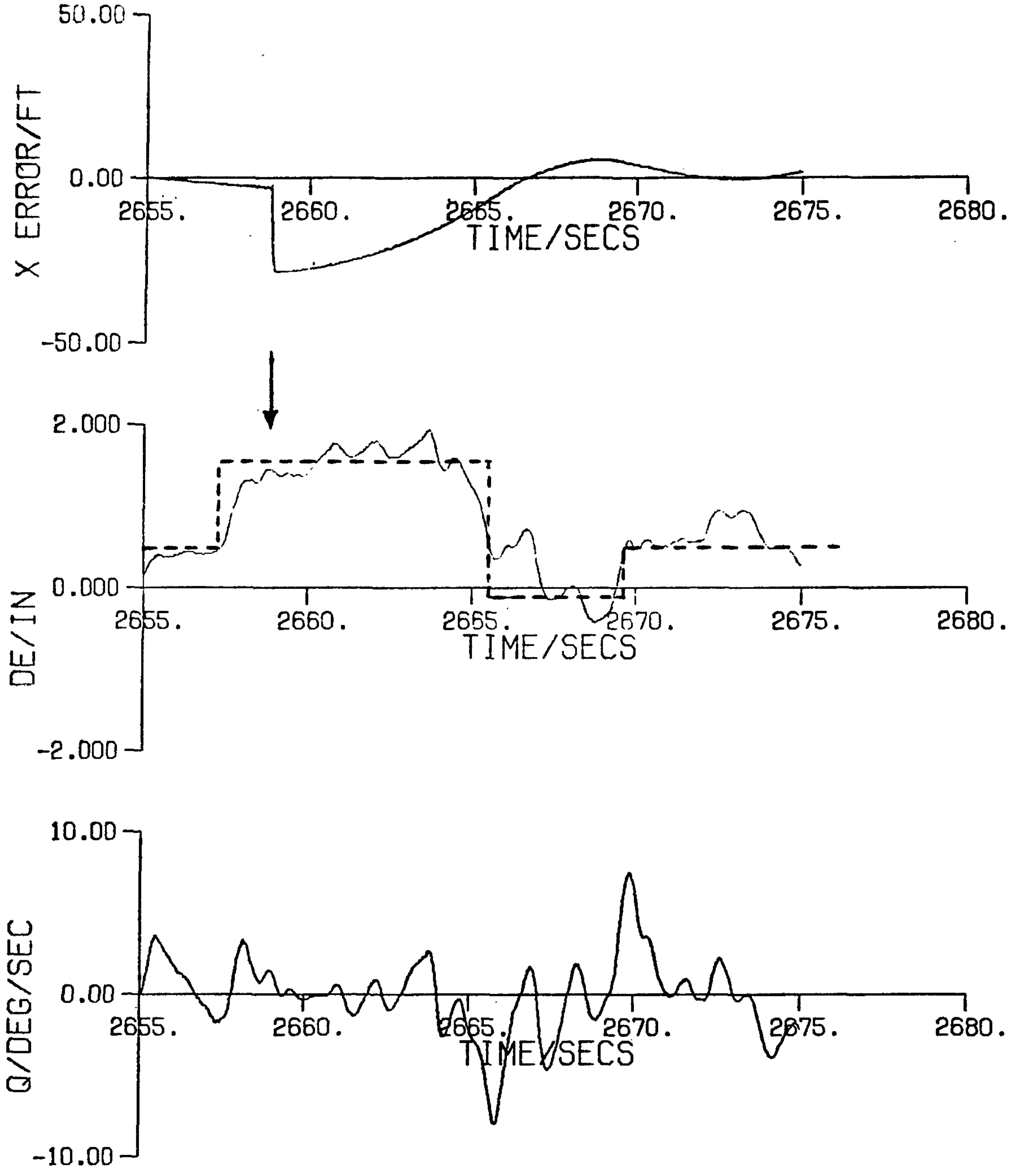


FIG. 8 X-22A FLIGHT DATA HISTORIES
ID:206B
F.C.:624

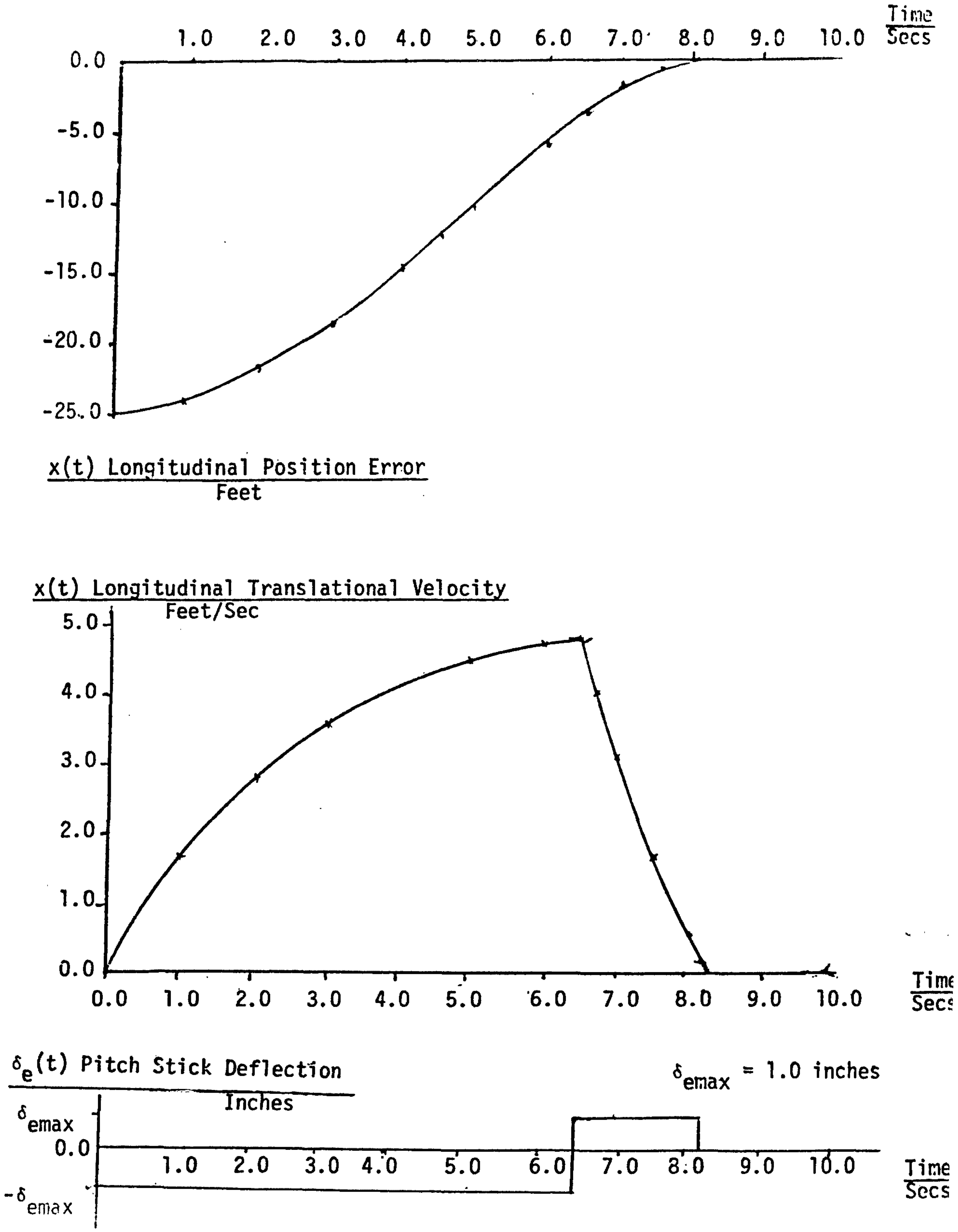


FIG. 9 TRANSIENT SOLUTION FOR IDENTIFIED BANG-BANG CONTROL MODEL

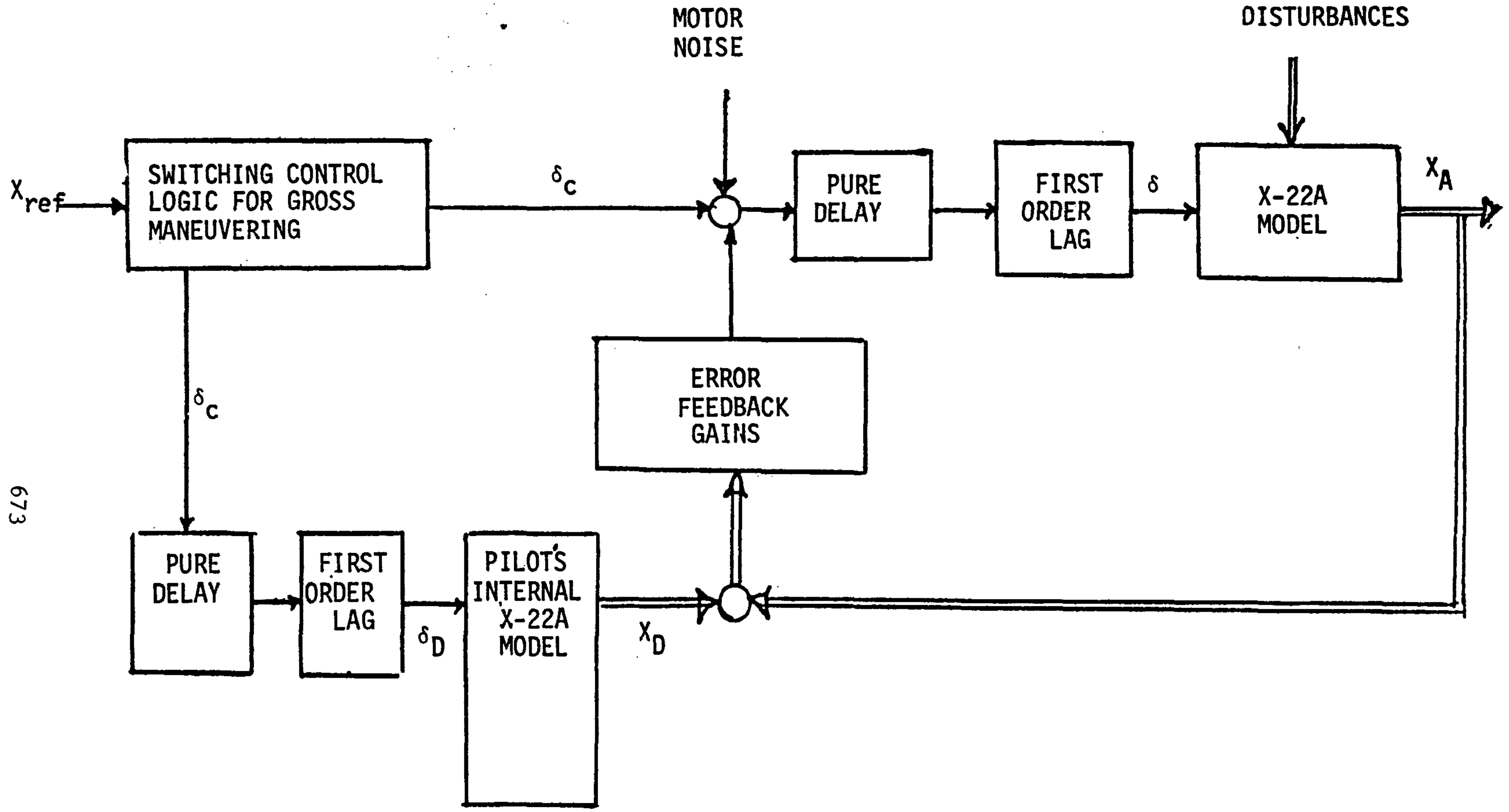


FIG. 10 PROPOSED PILOT MODEL

Pilots Internal Desired Trajectory Model

$$\tau \dot{\delta}_D(t) = - \delta_D(t) + \delta_C(t-t_o) \quad , \quad \delta_D(o) = 0$$

$$\dot{X}_D(t) = A\, X_D(t) + B\, \delta_D(t) \quad , \quad X_D(o) = 0$$

Pilot Output Model

$$\tau \dot{\delta}(t) = - \delta(t) + \delta_C(t-t_o) - K_p(X_D(t-t_o) - X_A(t-t_o))$$

$$+ m(t-t_o) \; , \quad \delta(o)$$

Vehicle Model

$$\dot{X}_A = A\, X_A(t) + B\delta(t) + CD(t) \quad , \quad X_A(o)$$

Measurement Equations for Parameter and State Estimation

$$X_m(t) = X_A(t) + n_X(t) + b_1$$

$$\delta_m(t) = \delta(t) + n_E(t) + b_2$$

Unknowns

δ_C, τ, t_o, K_p, $\delta(o)$, $X_A(o)$, b_1, b_2

Figure 11

Pilot Model Equations of Motion

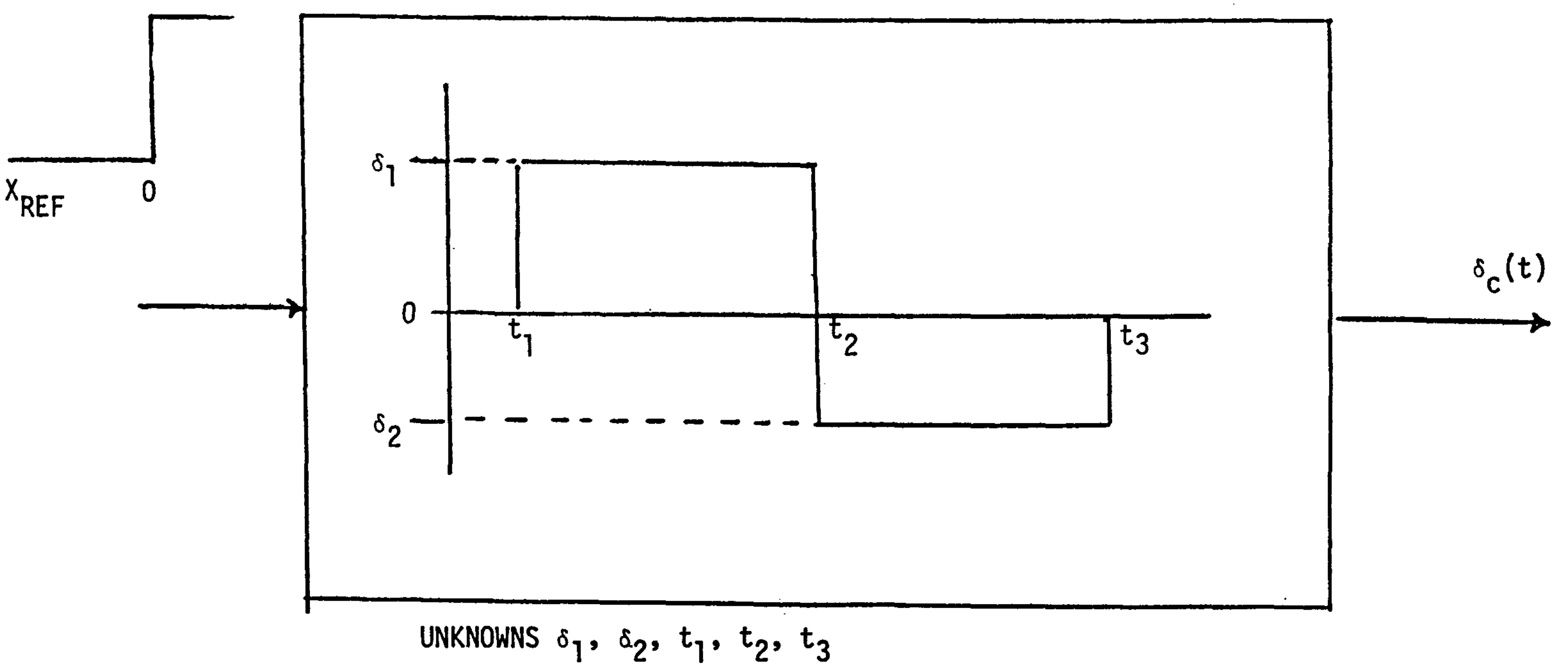


FIG. 12 PILOTS SWITCHING CONTROL

Flight Measurements of Pilot Performance for UH-60 Simulation Validation Assessment. (Progress Report)

William B. Cleveland
NASA-Ames Research Center
Moffett Field, California

Robert K. Heffley
Systems Technology Incorporated
Mountain View, California

David L. Key
Aeromechanics Laboratory
U.S. Army Research and Technology Laboratories (ADVRADCOM)
Ames Research Center
Moffett Field, California

ABSTRACT

A joint NASA/Army effort to develop a qualitative assessment of simulator fidelity is underway at the Ames Research Center. The approach is to first obtain a highly accurate mathematical model and then couple this with an appropriate visual, motion, and computation system. The vehicle being used for this study is the Army Sikorsky UH-60 Blackhawk helicopter. A complex math model has been obtained from Sikorsky and programmed for real time operation at the Ames Research Center. The model will be verified by comparing with flight data being obtained in a special flight test program performed by the Army engineering flight activity at Edwards Air Force Base in the period September 1981 through June 1982. As a corollary to the testing for the mathematical model parameter identification, this same highly instrumented UH-60 will also be used to obtain pilot performance data to use in the simulation fidelity validation process. Following analysis of the flight data these simulations will be performed at Ames and an assessment made of the range of validity.

Methods for Identifying Pilot Dynamics

by

William H. Levison
Bolt Beranek and Newman, Inc.
Cambridge, MA 02238

Proceedings of the
Workshop on Flight Testing to Identify Pilot Workload
and Pilot Dynamics

January 19-21, 1982
Edwards Air Force Base, CA

ABSTRACT

Procedures for identifying pilot dynamics are reviewed, with emphasis on frequency-response analysis of data obtained in well-controlled test environments. This paper discusses some theoretical constraints relevant to the computation of operator describing functions and remnant spectra; reviews advantages and limitations of frequency-domain techniques commonly used; and describes the use of sum-of-sines inputs to enhance the measurement environment. Procedures for identifying pilot-related parameters of the Optimal Control Model are reviewed.

INTRODUCTION

Techniques for measuring pilot response behavior in closed-loop control tasks have reached an advanced state of development over the past few decades. Proper application of this technology, combined with experimental design that maximizes measurement capabilities, can lead to reliable and accurate estimates of human operator behavior. On the other hand, ill-suited techniques or unfavorable measurement situations may yield grossly inaccurate estimates of response behavior.

Measurement of pilot response is usually undertaken to describe response behavior in a particular control situation, and, in many cases, to determine a general set of rules that allows one to predict pilot response in a variety of control tasks. The end product of the measurement/analysis procedure is often of a set of coefficients for a particular pilot/vehicle model.

Quantification of pilot response behavior in terms of model parameters may be done directly or through an intermediate stage of data reduction. The one-stage procedure leads directly from experimental time histories to model parameters; the two-stage procedure first transforms the information contained in the time histories to a format that is more readily interpretable in terms of operator response strategy. Model analysis is then performed on the transformed data.

The bulk of this paper focuses on the first stage of the two-stage analytical procedure; specifically, the transformation of experimental time histories to frequency-response measures that relate to the linear portion of the pilot's response strategy (the describing function) as well as to the stochastic portion of the pilot's control response (the power spectral density of the "pilot remnant"). These metrics are useful for testing a variety of models, including both "classical" frequency-response models [1] as well as the optimal control model for pilot/vehicle systems described in a companion paper [2].

Discussion is limited to situations in which pilot response behavior is representable as a noisy, time-invariant linear process. This restriction implies that (1) the controlled plant is representable as a time-invariant linear system, (2) the external forcing function(s) have stationary statistical properties, and (3) the pilot has been sufficiently motivated and trained to the stage where he has stabilized upon a response strategy appropriate to the control task.

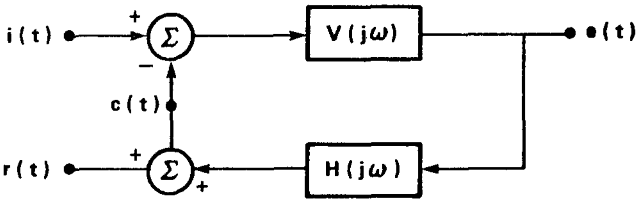
The remainder of this paper is organized into three major sections. The first section reviews certain theoretical aspects of frequency response analysis and points out the advantages of having external inputs available for measurement. Computational techniques appropriate to inputs continuous in frequency as well as discrete-frequency inputs (i.e., sum-of-sinusoids) are treated in the subsequent section. The final major section provides a brief description of a technique for identifying and statistically testing pilot-related parameters of the optimal control model.

THEORETICAL ANALYSIS

In this section we consider some of the limitations and sources of error associated with measurement of pilot response characteristics in closed-loop control tasks. Without presupposing any specific measurement techniques at this stage, we shall assume that cross correlations and/or cross power spectral densities (PSD) are obtainable between any two measured quantities. We show how analysis is influenced by system complexity (single- versus multi-variable systems), by the availability of the external input for measurement, and by the presence or absence of certain constraints on the computation of cross PSD.

Analysis of Single-Variable Systems

Let us consider the task of analyzing the single-variable feedback control system diagrammed in Figure 1. The objective is



i(t) = INPUT FORCING FUNCTION
e(t) = SYSTEM ERROR
c(t) = OPERATOR'S CONTROL INPUT
r(t) = OPERATOR REMNANT
H(jω) = OPERATOR DESCRIBING FUNCTION
V(jω) = CONTROLLED–ELEMENT TRANSFER FUNCTION

Figure 1. Diagram of Single Variable System

to estimate the pilot's transfer function H as well as the power spectral density of the "remnant" r(t). The pilot's remnant is assumed to be linearly independent of the external forcing function, and is not a directly measurable quantity. No assumptions are made with regard to the spectral content of the remnant.

The pilot's transfer characteristic is not, strictly speaking, a fixed quantity, because the pilot adjusts his strategy according to the nature of the control task. Accordingly, manual control research has adopted the term "describing function" to indicate the pilot's linear response strategy in a specific control situation. We shall adopt this terminology in the remainder of the paper.

The theoretical development presented below makes extensive use of the following properties of linear systems:

$$\begin{aligned} \Phi_{xy}(j\omega) &= H(j\omega) \cdot \Phi_{xx}(\omega) \\ \Phi^*_{xy}(j\omega) &= H^*(j\omega) \cdot \Phi_{xx}(\omega) \\ \Phi_{yy}(j\omega) &= |H(j\omega)|^2 \; \Phi_{xx}(\omega) \end{aligned} \tag{1}$$

where $\Phi_{xy}(j\omega)$ is the Fourier transform of the cross-correlation function relating input x(t) variable to output variable y(t), $H(j\omega)$ is the linear system function, and the operation (*) indicates complex conjugate. For economy of notation, the frequency argument $(j\omega)$ is omitted from the following analysis.

The following cross spectral density functions are derived from the above relationships:

Let $\Delta = 1 + HV$

$$\Phi_{ie} = \frac{V}{\Delta} \cdot (\Phi_{ii} - \Phi_{ir}) \tag{2}$$

$$\Phi_{re} = \frac{V}{\Delta} \cdot (\Phi^*_{ir} - \Phi_{rr}) \tag{3}$$

$$\Phi_{ic} = \frac{HV}{\Delta} \cdot (\Phi_{ii} + \frac{1}{HV} \Phi_{ir}) \tag{4}$$

$$\Phi_{rc} = \frac{1}{\Delta} \cdot (\Phi_{rr} + HV \cdot \Phi^*_{ir}) \tag{5}$$

$$\Phi_{ee} = |\frac{V}{\Delta}|^2 \cdot (\Phi_{ii} + \Phi_{rr} - [\phi_{ir} + \phi^*_{ir}]) \tag{6}$$

$$\Phi_{ec} = \left|\frac{HV}{\Delta}\right|^2 \cdot (\Phi_{ii} + \frac{\Phi_{rr}}{|HV|^2} + \frac{\Phi_{ir}}{HV} + \frac{\Phi^*_{ir}}{H^*V^*} \tag{7}$$

$$\Phi_{ec} = H \cdot \left|\frac{V}{\Delta}\right|^2 \cdot (\Phi_{ii} - \Phi^*_{ir} + \frac{1}{HV} [\Phi_{ir} - \Phi_{rr}]) \tag{8}$$

Because of the assumed linear independence between the external forcing function and operator remnant, the cross PSD term Φ_{ir} should be zero. Even when these signals are truly independent, however, limitations on the measurement interval may introduce an apparent correlation. As we show later, this apparent correlation can be minimized through appropriate experimental and analytical techniques.

The task of identifying pilot response characteristics is greatly facilitated when the external inputs are available for measurement, as is usually the case in ground-based simulation experiments. These signals are often absent for data obtained in-flight, however. In the following discussion we first summarize some of the problems of analyzing data bases in which external inputs are absent, and we then show how measurement accuracy can be improved when these inputs are available.

External Inputs Unavailable, Unconstrained Correlations

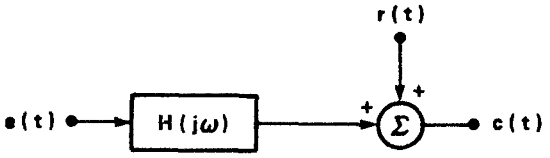
Let us consider the measurement situation, diagrammed in Figure 2a, in which only the system error and pilot's control input are available for measurement. The external input (e.g., simulated wind gust, target motion, etc.) is assumed unavailable. The "remnant" input r(t) is internal to the pilot and is therefore not directly measurable. We first consider measurement capabilities when no constraints are placed on the measured cross correlations.

Application of the properties of equation (1) to the system shown in Figure 2a yields the following relationships:

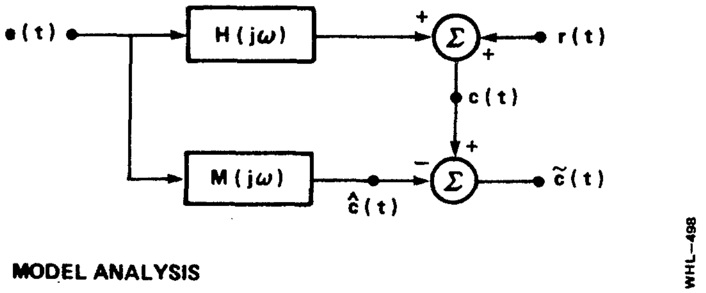
$$\begin{aligned} \Phi_{ec} &= H \cdot \Phi_{ee} + \Phi^*_{re} \\ \Phi_{cc} &= |H|^2_{ee} + H^*\Phi_{re} + H\Phi^*_{re} + \Phi_{rr} \end{aligned} \tag{9}$$

from which we derive an additional useful relationship:

$$\Phi_{cc} - \frac{|\Phi_{ec}|^2}{\Phi_{ee}} = \Phi_{rr} - \frac{|\Phi_{re}|^2}{\Phi_{ee}} \tag{10}$$



a) UNCONSTRAINED MEASUREMENTS



b) MODEL ANALYSIS

c(t) = MODELING ERROR
M(jω) = TRANSFER FUNCTION OF OPERATOR MODEL

Figure 2. Measurement Situation with External Input Unavailable

In the ideal measurement situation, there is no cross correlation between tracking error and pilot remnant, in which case $\Phi_{re} = 0$. Since r(t), and thus Φ_{re}, are not measurable, one has little choice but to assume ideal measurement conditions and to utilize the following estimation strategy based on the relationships shown in equations (9 and 10).

$$\hat{H} = \Phi_{ec}/\Phi_{ee}$$
$$\hat{\Phi}_{rr} = \Phi_{ec} - \frac{|\Phi_{ec}|^2}{|\Phi_{ee}|} \tag{11}$$

where the symbol (^) signifies an estimated quantity. In general, however, Φ_{re} is <u>not</u> zero, and the estimated pilot response variables are related to the measurable quantities as follows:

$$\hat{H} = H + \Phi^*_{re}/\Phi_{ee}$$

$$\hat{\Phi}_{rr} = \Phi_{re} - \frac{|\Phi_{re}|^2}{\Phi_{ee}} \qquad (12)$$

Using the relationships of equations (6-8), we obtain the following expression for the estimated describing function in terms of external signals driving the system:

$$\hat{H} = H \cdot \frac{1 - \frac{1}{HV} \frac{\Phi_{rr} - \Phi_{ir}}{\Phi_{ii} - \Phi^*_{ir}}}{1 + \frac{\Phi_{rr} - \Phi_{ir}}{\Phi_{ii} - \Phi^*_{ir}}}$$

$$(13)$$

$$\hat{H} = \begin{cases} H \text{ if } (\Phi_{rr} - \Phi^*_{ir}) << (\Phi_{ii} - \Phi_{ir}) \\ -\frac{1}{V} \text{ if } (\Phi_{rr} - \Phi^*_{ir}) \quad (\Phi_{ii} - \Phi_{ir}) \end{cases}$$

Thus, where pilot remnant is low, a good estimate to the pilot's describing function is obtained. Where remnant is relatively large, however, the estimate tends toward the negative inverse of the vehicle transfer function and is unrelated to pilot response strategy.

A similar situation arises with respect to the estimate of pilot remnant, which is related to external signals by

$$\hat{\Phi}_{rr} = \Phi_{rr} \cdot \frac{1 - \dfrac{\Phi_{ir}^{\;2}}{\Phi_{ii}\,\Phi_{rr}}}{1 + \dfrac{\Phi_{rr} - (\Phi_{ir} + \Phi^*_{ir})}{\Phi_{ii}}} \tag{14}$$

$$\Phi_{rr} = \begin{cases} \Phi_{rr} \text{ if } \Phi_{rr} << \Phi_{ii}, \Phi_{ir} << \Phi_{rr} \\ \Phi_{ii} \text{ if } \Phi_{rr} << \Phi_{ii}, \Phi_{ir} << \Phi_{rr} \end{cases}$$

Again, a reliable estimate is obtained when remnant is low; but for large remnant, the estimate of the pilot remnant spectrum approaches the spectrum of the external input and bears no relationship to pilot response behavior.

To summarize the theoretical analysis so far, when pilot response parameters are based only on measurements of system error and pilot control input, errors in estimating both the pilot describing function and remnant spectrum are likely if the total correlation between error and control is considered. Such errors arise largely from the failure to consider cause and effect. Specifically, even though pilot remnant is theoretically independent of the other external inputs to the system, a "backward correlation" between system error and remnant will generally be measured due to the response of the system to the pilot's input. This correlation distorts the estimates of pilot response parameters, sometimes to the point that the estimated describing function does not reflect a physically realizable system. As discussed below, improved estimates of pilot describing function and remnant are possible in high-remnant situations through measurement techniques that reduce the influence of the "backward correlation".

External Inputs Unavailable, Constrained Correlation

Model-matching techniques are often employed to impose desired cause and effect relationships on the analysis procedure when external inputs are not available for measurement. As shown in Figure 2b, one obtains the difference between the measured pilot input $\hat{c}(t)$ and the signal $\hat{c}(t)$ predicted by filtering the error signal through some model, defined by transfer function $M(j\omega)$. The model M is adjusted to minimize the matching error $\tilde{c}^2(t)$, where $\tilde{c}(t)$ is the difference between the actual and predicted control signals. The model M is then taken as the estimate of the pilot describing function.

The following spectral quantities are derived for $\tilde{c}$:

$$\Phi_{e\tilde{c}} = \tilde{M} \, \Phi_{ee} + \Phi^{*}_{re} \tag{15}$$

$$\Phi_{\tilde{c}\tilde{c}} = |\tilde{M}|^{2} \, \Phi_{ee} + \tilde{M}^{*}\Phi_{re} + \tilde{M}\Phi^{*}_{re} + \Phi_{rr} \tag{16}$$

where $\tilde{M}$ is the modeling error H-M. If the matching error is small, a good estimate of the remnant spectrum is given by

$$\hat{\Phi}_{rr} = \Phi_{\tilde{c}\tilde{c}} \tag{17}$$

even in the presence of sizeable cross correlations between error and remnant. If the matching error is not negligible, however, it will introduce errors into the estimated remnant spectrum as indicated in Equation (16).

A combination of non-negligible matching error and remnant-related tracking error may also introduce a bias in the estimate of the operator's describing function. Because the quantity $\tilde{M}^{*}\Phi_{re} + \tilde{M}\Phi^{*}_{re}$ (a real quantity) is not constrained to be non-negative, it may be possible to obtain minimum $\tilde{c}^{2}(t)$ (equivalently, minimum integral of the spectral density Φ_{cc}) for $M \neq H$.

The effects of the cross terms in $\Phi_{\tilde{c}\tilde{c}}$ can be minimized if the model M is constrained to be physically realizable (i.e., does not respond to inputs that have not yet occurred). Estimation reliability is further enhanced if the effective correlation time of the remnant signal (i.e., the reciprocal of the bandwidth) is less than the delay associated with the pilot's response [3,4].

By imposing the above constraints on his measurement procedure, Wingrove [4] has been able to obtain good results from in-flight and laboratory data in the absence of a measurable forcing function. The reader is referred to Wingrove for further discussion of these techniques.

Although minimizing the effects of the cross terms in equation (16) reduces one source of estimation error, this alone does not guarantee an accurate estimate of operator describing function and remnant spectrum. Implementation of a model-matching procedure imposes a particular structure on operator response behavior; to the extent that this structure is inadequate, estimation errors will be introduced.

External Inputs Available

The measurement situation is greatly enhanced when the external inputs are available for measurement: reliable estimates of both the human operator describing function and remnant spectrum may be obtained without cause-and-effect constraints placed on the cross correlations. Dividing Eq. (8) by Eq. (6), we obtain

$$\frac{\Phi_{ic}}{\Phi_{ie}} = H \, \frac{\Phi_{ii} + \frac{1}{HV}\Phi_{ir}}{\Phi_{ii} - \Phi_{ir}} \tag{18}$$

Since Φ_{ir} is theoretically zero (and, in practice, approaches zero for a sufficiently long measurement interval), the ratio of cross spectra shown here affords a good approximation to the pilot describing function. Therefore,

$$\begin{aligned} \hat{H} = \frac{\Phi_{ic}}{\Phi_{ie}} &= H \, \frac{1 + \frac{1}{HV}\Phi_{ir}/\Phi_{ii}}{1 - \Phi_{ir}/\Phi_{ii}} \\ &= H \quad \text{if} \quad \Phi_{ir} << \Phi_{ii} \end{aligned} \tag{19}$$

A good approximation to the closed-loop remnant spectrum is given by

$$\hat{\Phi}_{cc} = \Phi_{cc} - |\Phi_{ic}|^2/\Phi_{ii} \tag{20}$$

From equations 2-7, we show that, in terms of input- and remnant-correlated quantities, this expression is equivalent to

$$\begin{aligned} \hat{\Phi}_{cc_r} &= \left|\frac{1}{\Delta}\right|^2 \Phi_{rr} \quad \frac{|\Phi_{ir}|^2}{\Phi_{rr} \cdot \Phi_{ii}} \\ &\simeq \left|\frac{1}{\Delta}\right|^2 \Phi_{rr} \text{ if } |\Phi_{ir}|^2 << \Phi_{ii} \cdot \Phi_{rr} \end{aligned} \tag{21}$$

This closed-loop measure can be converted to an estimate of the open-loop (injected) remnant by scaling with $|1+HV|^2$.

As can be seen from the above development, use of the external forcing function in the analysis procedure avoids measurement errors introduced by large pilot remnant. Accurate estimates of pilot describing function and remnant may be obtained in the presence of large Φ_{rr}, provided the apparent input-remnant correlation Φ_{ir} is sufficiently small. A technique for minimizing Φ_{ir} is reviewed later in this paper.

Analysis of Multivariable Systems

Since considerations of length preclude a detailed treatment of multivariable systems, we present here a very brief summary of the analysis documented by Levison [3]. The (matrix) pilot describing function is estimated from the cross PSD's as follows:

$$\hat{\underline{H}} = -\underline{\Phi}_{ic} \cdot \underline{\Phi}_{ie}^{-1} \tag{22}$$

where H contains elements relating each "error" quantity to each control input, and the cross PSD's $\underline{\Phi}_{ic}$ and $\underline{\Phi}_{ie}$ relate each input to each control and each error variable, respectively.

In order for this computation to be valid, $\underline{\Phi}_{ie}$ must be invertible. This imposes the following rather severe restrictions: (1) there must be as many external forcing functions as there are error (i.e., display) variables; (2) these forcing functions must be, or be transformations of, linearly independent processes, and (3) the output variables must be differentially influenced by the forcing functions.

Even when these conditions are met, one cannot in general compute the transfer element H_{jk} from only the PSD's involving the jth control variable and the kth display variable; the matrix operation shown above must be carried out to yield readily interpretable measures of pilot performance. Sample multivariable control systems are analyzed in Reference 3 to illustrate this point. Also derived in that reference is a (rather complex) expression for the vector remnant process $\underline{\Phi}_{rr}$.

COMPUTATIONAL TECHNIQUES

Issues related to frequency-response analysis techniques are reviewed in this section. We briefly review certain important aspects of Fourier-series analysis, and we then consider continuous-frequency and sum-of-sinusoids measurement techniques.

Fourier Analysis

The advent of computationally-efficient fast-Fourier transform techniques [5] has greatly facilitated analysis of data obtained in manual control experiments, and the existence of these techniques has encouraged the design of experiments in which steady-state frequency-domain analysis is applicable.

Application of Fourier series methods assumes that a time series of finite length may be treated as if it were periodic. (We do not necessarily assume that the data are actually periodic, but rather that we can obtain the desired information using analysis procedures designed for periodic signals.)

If a signal is periodic in time T, we define a fundamental or "base" frequency as $\omega_o = 2\pi/T$, and we note that all frequency components of such a signal must be harmonics of this base frequency. The cross PSD is given as

$$\Phi_{xy} = \overline{Y(k)X^*(k)} \tag{23}$$

where k is an integer, X(k) and Y(k) are Fourier coefficients defined at frequency $k\omega_o$, and the overstrike indicates statistical expectation. Fourier coefficients and spectral quantities are defined for both positive and negative frequencies (i.e., positive and negative values of "k").

The quantity Φ_{xy} is most properly termed a "correlogram", rather than a "spectrum" when analyzing a finite-time sample of a random process. Although this quantity is often computed to obtain an estimate of the spectrum of a random process, the standard deviation of the error associated with such an estimate is large -- on the order of the expected value of the estimate [5]. The effects of this error, however, can be considerably reduced through appropriate averaging techniques. Additional improvement in estimation accuracy can be obtained if inputs are constructed as sums of sinusoids, in which case the input correlated portion of the signal is truly periodic.

In the remainder of this section we consider the analysis of a single-variable control task as diagrammed in Figure 1. We first consider inputs that are continuous in frequency, then we show the computational advantages of using sum-of-sines inputs.

Inputs Continuous in Frequency

Assume that the tracking input i(t) is continuous in frequency in the regions where frequency-response measures are desired. From the relationships of Eqs. (19) and (20), plus the property of Eq.(23), we derive the following describing function and remnant estimates in terms of measurable quantities:

$$\hat{H} = \frac{\Phi_{ic}}{\Phi_{ie}} = \frac{\overline{CI^*}}{\overline{EI^*}}$$

$$\hat{\Phi}_{c_{rr}} = \Phi_{cc} - \frac{|\Phi_{ic}|^2}{\Phi_{ii}} = \overline{|C|^2} - \frac{|\overline{CI^*}|}{\overline{|I|^2}} \qquad (24)$$

where I, E, and C are the Fourier coefficients of the input, error, and control signals, respectively, at a specific frequency $k\omega_o$. Φ_{cc_r} denotes the closed-loop remnant spectrum; i.e., the portion of the control power related to the remnant input. (For notational convenience, we shall omit the frequency argument for Fourier coefficients.)

These estimates are related to system variables as follows:

$$\hat{H} = H \frac{1 + \frac{1}{HV} \cdot \frac{\overline{RI^*}}{|I|^2}}{1 - \frac{\overline{RI^*}}{|I|^2}}$$

$$(25)$$

$$\Phi_{cc_r} = \frac{1}{|\Delta|^2} \left[\overline{|R|^2} - \frac{\overline{RI^*}^{\,2}}{\overline{|I|^2}} \right]$$

Thus, as the number of measurements available for averaging increases, the quantity $\overline{R}I^*$ tends toward zero (because of the presumed lack of linear correlation between the external input and operator remnant process), and the estimates approach their true values. The quantities H and V are assumed constant and are therefore extracted form the averaging process. If these quantities are not constant, then an average H would be estimated.

Averaging must be performed on the individual cross spectral quantities, rather than on estimates of H and Φ_{cc_r} obtained from individual sets of measurements, because one cannot distinguish between remnant-related and input-correlated signal power from single measurements of I, E, and C. Averaging may be done either across frequency (i.e., average measurements obtained from neighboring frequencies), across an ensemble (average measurements at a single frequency obtained from repeated trials of the experiment), or a combination of both. In any case, we must make the assumption that the system function H is constant across the frequency band and from trial to trial. This constraint requires that the frequency band across which averages are taken be relatively small. (In practice, this author has used bandwidth of about 1/4 octave for performing within-trial averaging of spectral quantities.)

Because the quantities to be averaged are cross-spectral measures and not the time signals or Fourier coefficients, one may use different inputs from trial to trial. Statistics of the input signals should be kept constant, however, to minimize trial to trial deviation in the pilot's response strategy.

The number of data points required to obtain a reliable estimate of H decreases with the relative amount of remnant present. Equation (25) shows that we get a "perfect" estimate of H if zero remnant is present. If substantial remnant is present, the measurement situation can be aided by concentrating input power in selected frequency bands and by averaging across repeated trials by the same subject and across subjects.

The task of obtaining reliable estimates of remnant power appears to be substantially more difficult than that of estimating the describing function. If the remnant power density is very small (which it typically is for control response at low frequencies under ideal tracking conditions), a numerical problem may arise. As Equation (24) shows, remnant power is computed as the difference between total and correlated power. If these estimates are nearly equal, precision may be lost due to roundoff errors.

Errors in estimating remnant also arise from the measured (apparent) correlation between the remnant and the input variable. Ideally, sufficient measurements would be averaged so that the effect of the quantity $\overline{RI^*}$ becomes negligible. This is usually not a severe problem for high-frequency measurements, where the averaging window (say, 1/4 octave) typically includes a substantial number of frequency intervals.* A more serious problem exists in obtaining accurate estimates of remnant at low frequencies, where the number of frequency intervals per window is smaller.

As described below, measurement inaccuracies arising from the cross-spectral quantity $\overline{RI^*}$ can be substantially reduced by the use of an external forcing function constructed as a sum of sinusoids.

Sum-of-Sines Inputs

An appropriate combination of sinusoids can provide a tracking input that appears to be stochastic to the subject but which overcomes many of the measurement difficulties associated with inputs that are continuous in frequency. This type of input allows us to model (and otherwise interpret) pilot behavior as a response to an unpredictable input, while providing the computational benefits of the highly structured sum of sinusoids.

If each component frequency is harmonically related to the reciprocal of the run length, and if each component sinusoid is generated without distortion, then Fourier analysis of the input signal will contain power only at the nominal input frequencies. Since there is no input power at non-input frequencies, any measurement of power at non-input frequencies (provided it is above irreducible system noise) is by definition related to operator "remnant". In order to estimate the remnant power density in the vicinity of a nominal input frequency, we simply average the power measurements over a frequency band on either side of (but not including) the input frequency. Estimates of Φ_{cc_r} can be averaged both across subjects and across trials for the same subject in order to improve accuracy.

With a sum-of-sines input, no averaging procedure is required to obtain estimates of the controller's describing function, and the expression of Equation (24) reduces to

---

* The number of "frequency intervals" contained in a frequency range $\Delta\omega$ is equal to $\Delta\omega/\omega_0$. Thus, estimation accuracy can be enhanced by increasing the run length, which decreases the base frequency and provides more frequency-response measurements for a given frequency range.

$$\hat{H} = CI^*/EI^* = C/E \qquad (26)$$

That is, the describing function is simply the ratio of the Fourier coefficients of the control and error signal at each input frequency.

Now, since the spectral measurement obtained at a given input frequency contains both remnant power as well as power truly correlated (i.e., linearly related) to the tracking input, computations of the Fourier coefficients C and/or E may be contaminated by operator remnant, introducing errors in the estimation of the operator's describing function. Typically, these errors are small, at least for laboratory tracking tasks. By using a sum-of-sines input, we concentrate the input-correlated power at a few selected frequencies. Since remnant is continuous in frequency (an assumption confirmed in numerous laboratory tracking studies), input-correlated power at a given input frequency is usually considerably greater than remnant power overlapping the same frequency.

An indication of the reliability of the describing function estimate at a given frequency can be obtained from a comparison of the power measured at that frequency to the neighboring remnant power. If the power at the input frequency is sufficiently greater than the surrounding remnant power (say, by 7 dB or more) for <u>both</u> error and control signals, estimates of C,E, and H may be considered reliable. Otherwise, the estimate of H should be considered unreliable and should not be used as a basis for modeling or otherwise characterizing human operator behavior.

In summary, use of a sum-of-sines input provides increased accuracy in the estimation of operator describing functions and operator remnant spectra. Describing functions can be obtained without averaging cross spectral quantities, and estimates of the remnant spectrum are free from errors introduced by the (apparent) cross correlation between input and remnant that is present when inputs are continuous in frequency. Comparison of spectral measures at input and non-input frequencies provides an indication of the reliability of the describing function estimate.

Use of sum-of-sines inputs necessarily restricts estimation of operator describing function to selected frequencies. In most control situations, however, one can obtain measurements at a sufficient number of frequencies to adequately characterize operator response behavior.

In order to provide maximum analytical power, sum-of-sines inputs designed for manual control studies should meet the following criteria:

a. The measurement interval should contain an integral number of cycles of each component sinusoid. (Equivalently, each component frequency should be harmonically related to the reciprocal of the run length.)

b. A sufficient number of sinusoids should be combined with random phasing so that the input is random-appearing to the subject.

c. The effective bandwidth of the input should be as wide as possible to maximize the range of valid measurements, but not so wide as to unrealistically affect tracking behavior.

d. Amplitudes of component sinusoids should be selected so that the input approximates some desired theoretical spectral density function.

e. Phase relationships should be readjusted from trial-to-trial (for a given subject) to prevent learning of the input.

The reader is referred to Levison [3] for elaboration of these criteria as well as for procedures for generating sum-of-sines inputs; for converting discrete spectral estimates to equivalent continuous power spectral densities, sampling and filtering analog signals; and for correcting for measurement delays.

Although the foregoing discussion has been limited to single-variable tracking tasks, sum-of-sines inputs may also be used to estimate the matrix transfer function when the operator is provided with more than one display variable. As noted earlier, there must be as many independent tracking inputs as there are displayed variables. Analysis and design procedures for tasks utilizing multiple sum-of-sines inputs are discussed in Reference [6].

IDENTIFICATION OF PILOT-RELATED MODEL PARAMETERS

The methods described above for reducing time histories to frequency-response measures may be considered the first step of a two-stage procedure for identifying pilot response parameters. In this section we review the second step: the derivation of pilot model parameters from the reduced data.

Discussion is limited to the "optimal control model" for the human operator in steady-state control tasks. Typical independent -- or "pilot-related" -- parameters to be identified are time delay, motor time constant, (equivalently, a "cost" penalty on rate-of-change of control), motor noise covariance, and an observation noise covariance for each perceptual input variable used by the operator. Readers unfamiliar with this model are directed to a companion paper [2]. Additional details on parameter identification are given in [7,8].

We first review the quasi-Newton (QN) minimization procedure recently developed for identifying model parameters, and then summarize a qualitative scheme for evaluating the significance of task-related changes in parameter values.

Review of the Quasi-Newton Identification Procedure

The Basic Minimization Procedure

Consider the task of adjusting model parameters to minimize a scalar matching error $J=\underline{e}'\underline{W}\underline{e}$, where each element e_i of the column vector $\underline{e}$ is the difference between the ith measured data point and the corresponding model prediction, and each element w_i of the diagonal matrix $\underline{W}$ is a weighting coefficient. In a given application, the matching error J will correspond to a particular choice of parameter values $\underline{p}$. The objective of the search procedure is to find a new parameter set $p+\Delta\underline{p}$ such that J is minimized.

To implement the search scheme, we initially assume that model predictions (and, therefore, prediction errors) vary linearly with model parameters. Thus, $\Delta\underline{e} = \underline{Q}'\Delta\underline{p}$, where

$$q(i,j) = \partial e_j / \partial p_i \qquad (27)$$

Solving for minimum J as a function of $\Delta\underline{p}$, we obtain

$$\Delta\underline{p} = -[\underline{QWQ}']^{-1}\underline{QWe} \qquad (28)$$

Now, since model input/output relationships are seldom totally linear, two or more iterations of the procedure are required until some convergence criteria are satisfied. In some cases, the parameter change computed as shown in Eq(28) will yield a scalar matching error greater than the starting value. Therefore, it is often useful to augment the minimization procedure described above with a line-search scheme to optimize the magnitude of $\Delta\underline{p}$.

Implementation for Manual Control Studies

Application of the QN method for analysis of human operator performance in continuous control tasks has been reported by Lancraft and Kleinman [9]. Described below is a revised implementation that has been used by this author in recent manual control studies.

Two criteria must be defined in order to apply the identification procedure: (1) a definition of a scalar matching error to be minimized by the QN scheme, and (2) convergence criteria to determine when the minimum modeling error has been approached sufficiently closely to justify termination of the minimization procedure.

Matching error is similar to that used by Lancraft and Kleinman:

$$J = \frac{1}{N} \sum_{i=1}^{N_1} \left(\frac{G_i - \hat{G}_i}{\sigma G_i} \right)^2 + \frac{1}{N} \sum_{i=1}^{N_2} \left(\frac{P_i - \hat{P}_i}{\sigma P_i} \right)^2$$

$$+ \frac{1}{N} \sum_{i=1}^{N_3} \left(\frac{R_i - \hat{R}_i}{\sigma R_i} \right)^2 + \frac{1}{N} \sum_{i=1}^{N_4} \left(\frac{S_i - \hat{S}_i}{\sigma S_i} \right)^2 \qquad (29)$$

where N is the number of valid measurements in the jth measurement group; G,P, are the gain (dB) and phase shift (degrees) of the ith describing function point to be matched; R is the corresponding control-stick "remnant" measurement (dB); and S is the ith variance score to be matched (units different for different tracking variables). σ indicates standard deviation of an experimental data point, and the symbol (^) indicates a model prediction.

Inclusion of the experimental deviations in the scalar modeling error allows each error component to be weighted inversely by the reliability of the data. To prevent the matching criterion from giving excessive weights to variables that have very low experimental variability, the following minimum standard deviations are imposed: 0.5 dB for magnitude and remnant, 3 degrees for phase, and 5% for the ensemble mean for variance scores.

Weighting inversely by standard deviation also converts each error term into a dimensionless number, thereby allowing accumulation of matching errors into a single metric. Thus, the matching error defined in Eq(29) approximates the average number of standard deviations of mismatch. A numerical score of J=4 reflects an average modeling error of 1 standard deviation (i.e., an average error of unity per measurement group).

The minimization procedure is terminated when the following conditions jointly obtain for two successive iterations: (1) reduction of the matching error by less than 0.5%, and (2) changes in all identified parameters by less than 2%.

A number of modifications have been made to the original implementation in order to improve computational efficiency. First, the search is performed on the logarithms of the parameters. This transformation modestly increases the degree of linearity between model parameters and model outputs, and it prevents the assignment of out-of-bounds (i.e., negative) values to parameters during the course of the search. Second, in order to minimize numerical difficulties with inversion of the expression QWQ', we omit from the search procedure (i.e., keep fixed), at a given iteration, any parameter having a negligible influence on the matching error. In addition, to reduce the chance of convergence to a local minimum appreciably removed from the global minimum, an individual parameter is allowed to undergo no more than a ten-fold increase or decrease from one iteration to the next.

Finally, a binary section scheme is employed to prevent divergence of the QN scheme due to nonlinear relationships between model inputs and outputs. If necessary, binary section is repeated until (1) matching error is reduced from one iteration to the next, or (2) until four attempts fail to reduce matching error, at which point the minimization scheme is terminated. Further details regarding implementation have been documented by Levison [7].

As is true with any numerical search procedure, the probability of convergence to a global minimum is enhanced by the selection of an initial set of model parameters that are close to the optimal set. The following rules for initializing model parameters appears to yield good results with the QN procedure: (1) cost of control rate such that motor time constant = 0.1 seconds; (2) time delay = 0.2 seconds; (3) observation noise covariance to achieve a noise/signal ratio of -20 dB for each perceptual variable assumed to be utilized by the operator; and (4) motor noise covariance to achieve a noise signal ratio of -50 dB, normalized with respect to control-rate variance.

Significance Testing

If independent model parameters are observed to vary across tasks, or to vary for a given pilot as a function of training, one may wish to test for the statistical significance of the observed parameter variations. One approach is to test the hypothesis that the various data sets can be modeled by the same set of model parameters. Failure to support this null hypothesis indicates that parameter differences are also significant.

A cross-comparison scheme has been developed and tested against data obtained in manual control studies. This scheme, which employs the empirical sensitivity test described below, provides a qualitative significance test on parameter differences obtained from pairs of experimental conditions.

Assume that model parameters have been identified from two data sets corresponding to, say, the "baseline" and "test" experimental conditions; our task is now to test the null hypothesis that a single set of model parameters provides a near-optimal match to the baseline and test data. To perform this test, we first identify the following three sets of pilot parameters: (1) the set that best matches the baseline data, (2) the set that best matches the test data, and (3) the set that provides the best joint match to the baseline and test data. For convenience, we shall refer to the parameters identified in step 3 as the "average parameter set".

We next compute the following four matching errors:

J(B,B) = matching error obtained from baseline data, using parameters identified from baseline data (i.e., best match to baseline data).
J(B,A) = matching error obtained from baseline data, using the average parameter set.
J(T,T) = best match to test data.
J(T,A) = matching error obtained from test data, using the average parameter set.

Finally, we compute the following "matching error ratios": J(B,A)/J(B,B); J(T,A)/J(T,T); and, if we wish to reduce the results to a single number, the average of these two error ratios. In a qualitative sense, the greater the matching error ratios, the more significant are the differences between the parameters identified for the baseline and test conditions. In addition to providing a collective test of the entire parameter set, this scheme may also be useful to test a single parameter or a subset of parameters.

As shown by Levison [7], a good approximation to the joint match to multiple data sets can sometimes be obtained by simply matching the average data. Thus, to obtain the "average parameter set", one would first obtain a point-by-point ensemble average of the (reduced) baseline and test data, and then identify parameters to match the average data set. This procedure is valid if the same task description applies to the two experimental conditions; i.e., if both tasks can be modeled identically except for quantitative differences in pilot-related parameters. Experiments designed to explore training effects, environmental stress, or interference from other concurrent tasks often meet this restriction.

The identification scheme reviewed here -- including the qualitative cross-comparison method of significance testing -- has been recently utilized to explore the effects of training on pilot-related model parameters [7,8].

CONCLUSIONS

A variety of methods exist for identifying pilot response characteristics in manual control tasks. Theoretical analysis and practical experience indicate that pilot response can be most accurately identified when the following conditions obtain: (a) steady-state task environment, (b) ability to measure the external forcing function(s), and (c) construction of the tracking input(s) as a sum of sinusoids. Pilot-related parameters of the optimal control model for the human operator can then be identified from time-domain and frequency-response measures computed from experimental time histories.

REFERENCES

1. McRuer, D.T., and Krendel, E.S., "Mathematical Models of Human Pilot Behavior", AGARD-AG-188, January 1974.

2. Levison, W.H., "The Optimal Control Model for the Human Operator: Theory, Validation, and Application", Proc. of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, Edwards Air Force Base, CA, January 1982.

3. Levison, W.H., "Measurement of Human Operator Response Behavior", BBN Report No. 4307, Bolt Beranek and Newman Inc., Cambridge, MA, February 1980.

4. Wingrove, R.C., "Comparison of Methods for Identifying Pilot Describing Functions from Closed-Loop Operating Records", NASA TN D-6235, Ames Research Center, Moffett Field, CA, March 1971.

5. Oppenheim, A.V., and Schafer, R.W., "Digital Signal Processing", Prentice-Hall, Inc., Englewood Cliffs, NJ, 1975.

6. Stapleford, R.L., Peters, R.A., and Alex, F., "Experiments and a Model for Pilot Dynamics with Visual and Motion Inputs", NASA CR-1325, May 1969.

7. Levison, W.H., "Effects of Whole-Body Motion Simulation on Flight Skill Development", BBN Report No. 4645, Bolt Beranek and Newman, Inc., Cambridge, MA, October 1981.

8. Levison, W.H., "A Quasi-Newton Procedure for Identifying Pilot-Related Parameters of the Optimal Control Model", Proc. of the Seventeenth Annual Conf. on Manual Control, Los Angeles, CA, June 1981.

9. Lancraft, R.E., and Kleinman, D.L., "On the Identification of Parameters in the Optimal Control Model", Proc. of the Fifteenth Annual Conference on Manual Control, Wright State University, Dayton, OH, March 1979.

THE AFTI/F-16 FLIGHT TEST PROGRAM AND OPPORTUNITIES TO EVALUATE PILOT-VEHICLE INTERFACE AND MISSION EFFECTIVENESS

by

Robert B. Crombie, Capt, USAF
Flight Dynamics Laboratory

Michael L. Frazier
Air Force Flight Test Center

The Advanced Fighter Technology (AFTI)/F-16 program is in response to today's European scenario, characterized by increased numbers of enemy targets both on the ground and in the air and an increasingly hostile air space surrounding these targets. This changing environment requires timely improvements in present USAF fighter lethality and survivability. The primary and continuing objective of the AFTI program, co-sponsored by the Air Force, NASA and Navy, is to provide for the development, integration, flight evaluation, and demonstration of emerging fighter technologies, and transition of the integrated technologies to future system applications. The AFTI Fighter Attack Technology (AFTI/F-16) program will develop, integrate and flight test a set of technologies to improve the survivability and weapon delivery accuracy of tactical fighters in air-to-air and air-to-ground attacks, through integration of advanced technologies into a single seat demonstrator vehicle which permits a realistic evaluation of technology benefits, penalties and overall mission effectiveness.

The AFTI/F-16 vehicle (depicted in Figure 1) has particular importance as a long life demonstrator aircraft with the flexibility, versatility, and capability in terms of performance and systems to serve as a future technology development testbed. A full-scale development F-16 aircraft is the test vehicle. Extensive modifications were made for installation of a sophisticated data instrumentation system, modified inlet with canards, new flight control system, and a dorsal fairing to accommodate the instrumentation equipment. Additional information on the AFTI/F-16 can be found in Reference 1.

The overall objective of the AFTI/F-16 Advanced Development Program is to demonstrate separately, and in combination, advanced fighter technologies to improve air-to-air (AA) and air-to-surface (AS) weapon delivery accuracy and survivability. These technologies include a Digital Flight Control System (DFCS), Automated Maneuvering Attack System (AMAS), pilot/vehicle interface (PVI) advancements, and advanced task-tailored control modes utilizing direct force control and weapon line pointing. Development, integration, and flight validation of these fighter attack technologies have been separated into DFCS and AMAS program phases.

The DFCS is a full-authority, triplex, digital fly-by-wire flight control system. The DFCS is mechanized to implement task-tailored manual control modes, including decoupled (six independent degrees of freedom or control-configured vehicle) flight control. Figure 2 shows that the pilot need only push a button to change the functions of cockpit controllers and displays. For the AMAS phase, the effective utilization of the advanced technologies requires the integration (coupling) of the fire and flight control functions. The integrated system will tie together a director fire control system, an advanced sensor-tracker, and the flight control system to provide precise automated weapon line control and weapons delivery. With the coupled system the azimuth and elevation fuselage pointing capability of the aircraft provides an expanded envelope of fire control solutions; i.e., an enlarged pipper. The pilot need only capture the target within the expanded pipper envelope and the fire control system will automatically command the flight control system to null aiming errors to assure a hit. This concept will profoundly influence fighter effectiveness in both AA and AS missions.

Pilot/vehicle interface advancements will be incorporated to provide crew station capabilities and environment commensurate with the increase in total vehicle capabilities provided by the other technologies in each phase. The DFCS phase will focus on core technology development. The technologies of prime interest will be manual flight path control, avionics integration, and advanced controllers and displays. In the AMAS phase the allocation of function between the pilot and vehicle will be redistributed as a result of the DFCS experience. Those tasks best performed by the machine will be automated. Technological advances in sensors, fire control modes, and weapons fusing will be integrated with the DFCS capabilities. Figure 3 shows a rough schedule of the AFTI/F-16 program and the relative time frame where technological transition will occur from DFCS to AMAS, and from the AFTI/F-16 program to current and future fighter aircraft.

An example of advanced technology integration and utilization is in the AMAS precision low altitude maneuvering attack scenario. The technologies involved in this scenario include:

1. Flight path control with full authority digital flight control.

2. Task automation with integrated flight and fire control and low altitude radar autopilot.

3. Advanced sensor-tracker with low drag FLIR and laser ranger installation.

4. Integrated avionics and weapons fusing.

5. Cockpit development including multi-purpose displays, wide field of view heads up display, helmet-mounted sight and voice command.

6. Weapons interface with pilot consent and auto-release.

These technologies together give the AFTI/F-16 the ability to more effectively attack ground targets as depicted in Figure 4. A low altitude radar autopilot allows survivable ingress and egress. AMAS automated air-to-surface bombing modes provide the capability for flexible target acquisition, precise tracking, automated ingress/attack steering, and automated weapon release for both low altitude, or standoff delivery direct, or high-g turning attacks.

Target acquisition of either preplanned or in-flight designated targets can be accomplished through use of the helmet-mounted sight (HMS) for wide field of coverage visual designations, the heads up display (HUD) for narrow field of coverage visual designations, or the radar ground map for night/all weather, standoff designations. Once target acquisition has been obtained the AFTI sensor/tracker with imaging FLIR and laser ranger can be used for track, ID, and precise position measurement of the target. The pilot can jink manually to avoid threats until engaging an automated attack mode which may follow, at the pilot's option, either a conventional wings-level or high-g turning weapon delivery trajectory with automatically computed weapon release. In this scenario the AFTI/F-16 frees the pilot to become a true weapons system manager, concentrating on target acquisition, attack planning and threat avoidance. Figure 5 shows that this attack profile is clearly more survivable than that of the basic F-16A. The AMAS profile is expected to give an order of magnitude increase in survivability without degrading weapon delivery accuracy.

PILOT-VEHICLE INTERFACE EVALUATION

The evaluation of pilot-vehicle interface (PVI) during the AFTI/F16 development and flight test program will indeed be challenging. The opportunity is there to tie together existing knowledge in several disciplines: flying qualities, pilot dynamics, pilot workload, automation, and mission effectiveness. During the first year of flight testing the emphasis will be on manual control of an aircraft having eight different task-tailored flight control modes. Advances in our understanding of flying qualities and pilot dynamics could be made. Also, new measures of pilot workload and mission effectiveness will be checked out. In the second year of flight testing the emphasis will shift to automated control of the aircraft to improve mission effectiveness and reduce pilot workload. Criteria in those areas are expected to be explored, validated and applied.

Presently, planning for the first year of flight testing is virtually complete. Several pilot-in-the-loop (closed loop) maneuvers are contained in the test plan. They are drawn from various test activities in support of different engineering disciplines, but are listed together in Table 1. Each run enumerated in the table corresponds to a particular test configuration and flight condition called out in the flight test plan. In general, the AFTI/F-16 vehicle will respond differently for each of these runs. The closed-loop maneuvers will be used, from the pilot-vehicle interface viewpoint, to (1) determine the closed loop dynamic characteristics of the AFTI/F-16 vehicle, (2) generate pilot subjective ratings and comments while performing various tasks, (3) quantitatively measure workload, and (4) measure task performance (mission effectiveness). Pilot technique or pilot dynamics may also be studied. This wealth of flight test data will allow, within available resources, a broad assessment of various criteria for flying qualities, workload, and mission performance. Taken as a whole, this broad assessment offers a unique opportunity to increase our understanding of pilot-vehicle interface (PVI) and how best to evaluate PVI in flight.

Flying Qualities

A major portion of the first year of the AFTI/F-16 flight test program will be the flying qualities evaluation of the manual control modes. There are in essence eight different AFTI/F-16 airplanes corresponding to the eight task-tailored manual control modes, each of which must demonstrate satisfactory flying qualities. Many of these control modes are unconventional (decoupled force or pointing). New flying qualities criteria for such modes have been explored (see Reference 2 where a simple frequency-response criterion is evaluated), but in general the latest flying qualities specifications lack flight-test validated evaluation criteria. The opportunity exists in the AFTI/F-16 flight test program to explore new flying qualities criteria in a mission-relevant context. The following paragraphs will describe how this may be done using the state-of-the-art in flight test techniques, instrumentation, and data analysis capabilities.

The time responses of the AFTI/F-16 vehicle to various cockpit controller inputs will be compared with predictions using "canned" maneuvers (such as doublets) where the pilot is not performing any real task with the vehicle. These response time histories will be used to help validate simulations of the AFTI/F-16 which then may be used to generate analytical flying qualities data to compare against suitable criteria. However, pilot-in-the-loop (closed loop) maneuvers listed in Table 1 must be relied upon to provide the pilot a task to perform so that he can give a valid pilot rating and good pilot comments (see Reference 3). The details of these closed loop maneuvers are contained in the appendixes. Broadly, the PVI evaluation maneuvers can be grouped, relative to their use in the evaluation of flying qualities, as follows:

a. Frequency Response Analysis Maneuvers. These are structured to excite the motion of the vehicle so that a fast-fourier transform of the resulting time histories will yield a good frequency-domain representation of the aircraft response to pilot control inputs. Flying qualities analysis of aircraft frequency response can then be made.. These maneuvers provide the pilot with a realistic tracking task to perform although he is asked to increase his concentration and gain to a level that artificially excites the motion of the aircraft. Tracking performance, read from gun camera film, has in the past not correlated with pilot ratings. This is because, in pilot dynamics terminology, the pilot has increased his remnant (noise). However, good pilot ratings and comments are generated. These maneuvers include handling qualities during (air-to-air) tracking (see also Reference 4) and air-to-surface gun and bomb tracking.

b. Artificial Test Maneuvers. In a sense all test maneuvers are artificial, but the weapons separation maneuvers are exceptional. These maneuvers, described in Appendix B, will tax the pilot's ability to attain specific flight conditions and unusual attitudes within precise tolerance goals over a predetermined weapons release point. Flying qualities, pilot dynamics, and workload studies will benefit from these artificially intense maneuvers.

c. Continuous Task Maneuvers. Formation flying, refueling, landing (Appendix A), and flying in turbulence (Appendix C) are real mission tasks that will demonstrate the flying qualities of the AFTI/F-16 vehicle. Since they are continuous (in terms of constant flight condition and pilot technique) these maneuvers would be perfect for evaluation of pilot dynamics if a crucial limitation can be overcome. This is the difficulty of measuring the actual command signal during flight to which the pilot is responding. For example, how can we measure distance from the lead airplane (formation flying) or the distance to the refueling boom? In some cases the command signal can be picked off the heads-up display or otherwise inferred.

d. Evaluation Maneuvers (Appendix D). These are intended to duplicate real combat scenarios. Each project pilot will be allowed to use his own technique based on his experience with the unique task-tailored flight control modes and standard F-16 fire control system. Flying qualities criteria developed in previous tests or analysis can be related to mission effectiveness and pilot workload during these maneuvers both subjectively and quantitatively. The last two sections of this paper will describe how that may be done. Pilot dynamics measured during these discrete operational maneuvers may also provide insight into how different pilots perform real combat tasks. However, pilot training and experience will be limited by constraints in simulator and flight time. This is a limitation in any advanced development program.

Instrumentation on the AFTI/F-16 is described in Appendix E. The vehicle will be heavily instrumented for many different engineering disciplines. A partial list of instrumentation parameters to be measured during flying qualities flight tests is given in Table 2. Data processing will be accomplished on both a real-time and post-flight basis. Since the projected flight rate of the AFTI/F-16 will be three flights per week, flight-to-flight decisions will be based on real-time engineering units strip charts, crossplots, and tabulations displayed in the NASA control room. Post-flight data processing consists of more extensive second-generation plots and listings as well as third generation processing performed by AFFTC engineers. Third-generation programs will be used for fast fourier transform analyses (FRA; see Reference 5), parameter identification (MMLE3; see Reference 6), and as prediction and standardization routines (EASY5; see Reference 7).

Pilot Workload

During the first year of the AFTI/F-16 flight test program physiological data will be recorded and analyzed to help answer questions about the pilot workload associated with advanced fighter aircraft technologies. Various subjective rating techniques will supplement physiological workload analyses. Two new subjective techniques to be used on the AFTI/F-16 project, System Operability Measurement Algorithm (SOMA) and Subjective Workload Assessment Technique (SWAT) are very adequately covered in References 8 and 9 in this proceedings. The recording and analysis of pilot physiological data will be performed in collaboration with the Air Force School of Aerospace Medicine (AFSAM), Air Force Flight Test Center (AFFTC) and NASA. For details on the development of this physiological measurement technique see Reference 10 in this proceedings. EKG is the primary physiological parameter to be measured during this phase. From the EKG data, heart rate (an variance) and R-wave amplitude (and variance) will provide insights into pilot workload. Respiration rate is also being investigated as an additional measurement parameter.

Figure 6 illustrates the development of the physiological workload measurement techniques. Flights with physiological monitoring will begin one month after first flight. Months two through four will concentrate on developing procedure, shaking out equipment, developing workload criteria and cross-validating physiological data collected earlier during simulator and centrifuge experiments (see References 10, 11 and 12). Months four through eight will consist of answering "simple" questions about the workload associated with technologies in a mission context (e.g. does a decoupled mode cause less pilot workload during fine tracking than the normal mode?). Months 8 through 12 will involve answering complex issues about technologies, making recommendations about allocation of function between pilot and machine, in terms of automation, for Phase II, and reporting findings.

Both micro and macro analyses of pilot workload will be conducted on flights of interest. Micro analyses will be made on selected maneuvers to determine the workload associated with a given technology within a given maneuver or mission segment. Maneuvers of high interest are listed in Table 1. Also, micro workload analysis will be made on an exception basis, i.e. where an unexpectedly high workload situation occurs. Macro analysis will be made to determine the fatigue incurred during that total flight.

Due to the small number of pilots (6) and the effects of learning in a complex aircraft, there are some minimum criteria for examining maneuvers of interest for statistical comparisons:

1. All pilots must fly the maneuver.

2. Each pilot repeats the maneuver three times.

3. Learning effects are stabilized or at least identified.

Table 3 illustrates the interaction of advanced technologies and flight test activities. Advanced technologies are advertised to enhance the pilot's ability to manage aircraft systems (i.e. control of aircraft, weapons system management and support systems management). Flight test activities include flying qualities, weapons separation and DFCS evaluations. Maneuvers within each flight test activity are listed in the appendixes. The desirability or goodness of advanced technologies will be determined by analyzing pilot workload and task performance for the appropriate flight test activity.

MISSION EFFECTIVENESS EVALUATION

Complementary to the evaluation of Pilot-Vehicle Interface (workload or flying qualities) is the measurement of task performance. Presently, few attempts have been made in flight to quantitatively measure and relate pilot workload to task performance during the same maneuver. Most of these attempts were helicopter flight tests (e.g. References 13 and 14). Fixed wing experience at AFFTC so far (Reference 15) is that task performance measures (i.e. rms pipper error statistics) recorded during Handling Qualities During Tracking Maneuvers do not correlate with airplane handling qualities as assessed by pilots. It is clear that better ways to measure and explain task performance are necessary to adequately evaluate Pilot-Vehicle Interface during the AFTI/F-16 flight test program.

Mission effectiveness will be the bottom - line measure of task performance during the AFTI/F-16 program. The tasks, in this case, represent the two tasks inherent in any combat mission: killing the target and surviving enemy threats. Mission effectiveness analysis then will quantify these two bottom - line measures of combat mission effectiveness achieved by the pilot and aircraft on a given combat maneuver as:

a. Weapon delivery accuracy (probability of killing air or surface targets with certain weapons) and

b. Survivability (probability of surviving certain ground-based defenses).

These are expected to provide a meaningful comparison of the relative usefulness of the AFTI/F-16 technologies to perform the intended combat missions. Since the two tasks of killing the target and surviving are often in conflict with each other, they will demand the pilot's total attention and effort and so should provide a sound basis for validating in-flight measures of workload as well as for demonstrating the mission relevance of advanced flying qualities criteria. New insight into what may be the primary contributors to weapon delivery accuracy and survivability measures will also be gained for each mission segment so that a preliminary assessment of the feasibility of a performance or workload specification may be made.

MISSION EFFECTIVENESS ANALYSIS The basic approach to mission effectiveness analysis on the AFTI/F-16 program is to use the General Dynamics simulator as the primary source of quantitative mission effectiveness data for meeting contractual mission effectiveness specifications.

Simulations are planned to compare the DFCS-phase AFTI/F-16 with the baseline F-16A aircraft and with the fully-integrated AMAS-phase AFTI/F-16 using common mission scenarios. Proposed data collection and analysis details for the DFCS Mission Effectiveness Simulations are contained in Appendix F. Flight testing will be done to verify simulation results using the proposed DFCS evaluation maneuvers described in Appendix D. Flight test and simulator data will be processed through the same mission effectiveness computer programs by General Dynamics to provide this verification. Subjective pilot opinion and physiological wokload data will also be used to independently validate simulation fidelity.

COMPARISON WITH PILOT WORKLOAD We expect that evaluating performance in two primary tasks (hitting the target and surviving) should allow a good comparison with physiological and subjective measures of workload because the two tasks normally saturate the pilot's cognitive abilities. Success by others with secondary task measures of workload also gives us confidence that dual mission effectiveness measures will relate with pilot workload. For each flight test maneuver the pilot will need to be properly instructed and motivated as if in combat, that is, to be as elusive as possible for survival while placing the pipper on the target long enough to assure a kill. After each flight, subjective pilot ratings and comments will be taken. Flight data will be processed by AFFTC, AFSAM, and General Dynamics and compared. Each pilot's individual physiological variability using a "control" maneuver (e.g. takeoff). These comparisons will not only validate workload measures but may also lead us to promisinig mission effectiveness criteria that may eventually be used as a performance or workload specification. The following steps show how this may take place:

1. Establish realisitc mission scenarios, initial conditions and measurement and data reduction procedures for collecting the necessary data. Determine utility simulation versus flight test.
2. Develop ways to normalize data for differences among pilots and their various skill levels and thecniques.
3. Relate flying qualities criteria to workload and mission effectiveness criteria. Recommend profitable avenues of research.
    a. DFCS-Phase of AFTI/F-16 program emphasizes manual control of flight where flying qualities criteria and specifications apply. Levels of flying qualities will be related to workload and mission effectivensss criteria.
    b. AMAS-Phase of AFTI/F-16 emphasizes automated control of flight where workload and mission effectivensss criteria (developed in the previous phase) apply instead of flying qualities. Candidate specification criteria for workload or mission effectiveness will be recommended for futher research.

MISSION EFFECTIVENESS CRITERIA Calculations of the probability of target kill or probability of survival require specialized computer programs (as in Reference 17). A simpler approach would be to develop mission effectiveness criteria which could be measured directly during flight test. Such criteria would need to be proven to be components of either probability of kill or survival. The AFTI/F-16 program will provide an opportunity to examine the components of the two mission effectiveness measure, accuracy and survivability for various mission tasks as described below.

1. Accuracy. Listed in Table 4 are some of the primary contributors to hitting the target successfully, assuming constant weapon characteristics. The primary contributor ought to be some measure of fine tracking ability which efficiently allows the pilot to attain precise weapon release conditions. A secondary contributor to accuracy is some measre of gross acquisition ability. These hypotheses will be validated through correlation of flight data with probability of target kill calculations.

2. Survivablity. For contstant threat characterisitic, probability of survival should be proportional to some measure of elusiveness. Possible measures of elusivenss are listed in Table 5. These primarily

apply to ground-based threats. For threats from enemy aircraft a primary contributor to vulnerability is loss of energy (airspeed or altitude). Secondary contributors to survivability are the primary contributors to weapon delivery accuracy since the ability to quickly kill the target reduces the time exposed to hostile enemy action.

SUMMARY

The AFTI/F-16 flight test program will offer many opportunities for the Air Force, NASA, and Navy to evaluate pilot-vehicle interface and mission effectivenss. The AFTI/F-16 vehicle contains an advanced systems capability that will enhance the pilots ability to control the aircraft. Automated modes will configure these systems to help the pilot and vehicle to better perform realistic combat mission tasks. The first year's flight test plan has many pilot-in-the-loop maneuvers that will permit the evaluation of flying qualities, pilot workload, weapon delivery accuracy, and survivability. New criteria are expected to be developed or validated in each of these areas. The results, if adequate resources are applied, should advance our capability to test and evaluate pilot-vehicle interface and mission effectiveness.

REFERENCES

1. Gill, Robert A. and St. Sauver, Charles L., Captain, USAF, Advanced Fighter Technology Integration - AFTI/F-16 Test Program Overview, AIAA-81-2352, First Flight Testing Conference, Las Vegas, Nevada, November 11-13, 1981.

2. Hoh, Roger H., et al, Development of Handling Quality Criteria for Aircraft With Independent Control of Six Degrees of Freedom, AFWAL-TR-81-3027, Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio, April 1981.

3. Twisdale, Thomas R., Preflight Briefing and Postflight Debriefing Outlines for Handling Qualities Testing and Evaluation, Flight Test Technology Branch Office Memo, February 1977.

4. Twisdale, Thomas R., and Franklin, David L., Captain, USAF, Tracking Test Techniques for Handling Qualities Evaluation, AFFTC-TD-75-1, Air Force Flight Test Center, Edwards AFB, California, May 1975.

5. Twisdale, Thomas R., and Ashurst, Tice A, Jr., System Identification From Tracking (SIFT), A New Technique for Handling Qualities Test and Evaluation (Initial Report), AFFTC-TR-77-27 Air Force Flight Test Center, Edwards AFB, California, November 1977.

6. Maine, Richard E. and Iliff, Kenneth W., User's Manual for MMLE3, a General FORTRAN Program for Maximum Likelihood Parameter Estimation, NASA Technical Paper 1563, November 1980.

7. Burroughs, J.D., Fath, A.F., and Lloyd, A.J.P., Environmental Control (ECS) Transient Analysis, AFFDL-TR-77-102, Air Force Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio, October 1977.

8. Williams, Kendrick N., "Assessment of Reliability and Validity for the Systems Operability Measurement Algorithm (SCMA), *Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics*, AFFTC-TR-82-5, Air Force Flight Test Center, Edwards AFB, California, (to be published).

9. Reid, Gary., "Subjective Workload Assessment Technique", *Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics*, AFFTC-TR-82-5, Air Force Flight Test Center, Edwards AFB, California, (to be published).

10. Reader, D.C., Wing Commander, RAF, "Physiological and Performance Parameters as Indices of Pilot Workload - An Analysis of Data from the AFTI/F-16 Project," *Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics,*AFFTC-TR-82-5 Air Force Flight Test Center, Edwards AFB, California, (to be published).

11. Repperger, D.W., Rogers, D.B., Van Patter, R.E. and Frazier, J., "A Study of Task Difficulty With a Subjective Rating Scale," *Proceedings Of the Workshop of Flight Testing to Identify Pilot Workload and Pilot Dynamics*, AFFTC-TR-82-5, Air Force Flight Test Center, Edwards AFB, California, (to be published).

12. Popplow, James R., Major, CAF, "Some Physiological Changes in Lateral (GY) Acceleration", *Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics*, AFFTC-TR-82-5, Air Force Flight Test Center, Edwards AFC, California (to be published).

13. Winn, Albert L. and Lewis, Richard B., II, *Pilot Workload During Instrument Flight*, The 30th Annual National Forum of the American Helicopter Society, Washington, D.C., May 1974.

14, Van De Graaff, R.C., "NLR Research on Pilot Dynamics and Workload," *Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics*, AFFTC-TR-82-5, Air Force Flight Test Center, Edwards AFB, California,(to be published).

15. Twisdale, Thomas R., and Fanklin, David L., Captain USAF, *Tracking Test Techniques for Handling Qualities Evaluation (ADDENDUM)* AFFTC-TD-75-1 (ADDENDUM), Air Force Flight Test Center, Edwards AFB, California, May 1980.

16. Nardi, Louis U., Kawana, Harry Y., and Greek, David C., *Ride Qualities Criteria Validation/Pilot Performance Study - Flight Test Results*, NASA Contractor Report 144885, September 1979.

17. Bell, Robert A., et al, *Integrated Flight/Fire Control System (Firefly II/A-10)*, AFWAL-TR-80-3002, Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio, January 1980.

TABLE 1

PVI EVALUATION MANEUVERS

DEVELOPMENT MANEUVERS	NO. RUNS
HANDLING QUALITIES DURING TRACKING	21
AIR-TO-SURFACE GUN TRACKING	6
AIR-TO-SURFACE BOMB TRACKING	10
FORMATION FLYING	4
AERIAL REFUELING	9
ILS APPROACH TO LANDING	2
RESPONSE TO TURBULENCE	10
WEAPONS SEPARATION	27

EVALUATION MANEUVERS	NO. RUNS
AIR-TO-SURFACE BOMBING TECHNIQUES - CONTINUOUSLY COMPUTED IMPACT POINT - DIVE TOSS - CONTINUOUSLY COMPUTED RELEASE POINT - LOW ALTITUDE DROGUE DELIVERY	44
AIR-TO-SURFACE STRAFE	39
AIR-TO-AIR GUNNERY	45

TABLE 2

FLYING QUALITIES INSTRUMENTATION LIST

Instrumentation parameters to be used during the Flying Qualities and DFCS Development flight tests are as follows:

PARAMETER

Sideslip - Delta Pressure
RAM Air Temp
Landing Gear Handle Pos.
Landing Gear Handle Pos.
Alt. Flap Switch
Alt. Flap Switch
Air Refuel Door Pos.
Air Refuel Door Pos.
IFFC Switch Position
MLG Weight-On-Wheels
Spin Chute Deploy
Spin Chute Jettison
Dogfight Switch
Missile O'Ride Switch
Weapon Release Button
Weapon Release
Trigger-1st Detent
Trigger-2nd Detent
AOA > 29°
Stall Warning
CADC Good
INS Good
Drag Modulation
Auto Stores
Norm. Accel - FLCC
Norm. Accel. - CG
Norm. Accel - FLCC Sel.
Lat Accel.
Lat. Accel - FLCC Sel.
Long. Accel. - FLCC
Pitch Angel
Roll Angle
Yaw Angle
Roll Rate
Pitch Rate
Yaw Rate
Forward Fuel Qty.
AFt. Fuel Qty.
L/H Wing Fuel Qty.
R/H Wing Fuel Qty.
Basic Fuel Flow
Total Fuel Flow
Fuel Temperature
LEF Hinge Mom. BL 62.15
LEF Hinge Mom. BL 96.68
LEF Hinge Mom. BL 125.1
LEF Hinge Mom. BL 145.4
L/H Flaperon HM (hinge Momen
R/H Flaperon HM
L/H Horizontal Tail HM
R/H Horizontal Tail HM
L/H Canard HM

PARAMETER

Mach Number
Altitude
Altitude
KCAS
KTAS
Ps
qc
Angle of Attack - NB
Angle of Attack - FLCC
L/H AOA Cone
R/H AOA Cone
AOA Side Mount
Sideslip Angle - NB
Sideslip Angle - FLCC
Sideslip - Lower Probe
R/H Canard HM
Rudder HM
N1-RPM
N2-RPM
Nozzle Area
L/H Wheel RPM
R/H Wheel RPM
Spin Chute Load

TABLE 3

DFCS PVI EVALUATION FACTORS

ADVANCED TECHNOLOGIES TO ENHANCE MANAGEMENT OF A/C SYSTEMS	FLIGHT TEST ACTIVITIES (TASKS)					
	FLYING QUALITIES (11 Maneuvers)		WEAPONS SEPARATION (9 Maneuvers)		DFCS EVALUATION (17 Maneuvers)	
CONTROL OF A/C						
6 DOF FLIGHT Coupled	Pilot Workload	Task Performance	Pilot Workload	Task Performance	Pilot Workload	Task Performance
Decoupled	"	"	"	"	"	"
UNCONVENTIONAL CONTROLLERS Coupled	"	"	"	"	"	"
Decoupled	"	"	"	"	"	"
MODE SELECTION	"	"	"	"	"	"
MANAGEMENT OF WEAPONS AND SUPPORT SYSTEMS	"	"	"	"	"	"
MULTIPURPOSE DISPLAYS	"	"	"	"	"	"
VOICE COMMAND	"	"	"	"	"	"
HELMET MOUNTED SIGHT	"	"	"	"	"	"
WIDE FIELD OF VIEW HUD	"	"	"	"	"	"

TABLE 4

ACCURACY MEASURES

P_{HIT}, P_{KILL} FOR CONSTANT WEAPON CHARACTERISTICS

PRIMARY CONTRIBUTOR IS FINE TRACKING ACCURACY

- o PIPPER ERRORS AT WEAPON RELEASE
- o TIME ON TARGET
- o RMS TRACKING ERROR
- o FLYING QUALITIES PARAMETERS

SECONDARY CONTRIBUTOR IS GROSS ACQUISITION

- o TIME TO STABILIZE PIPPER
- o AERODYNAMIC CAPABILITY

TABLE 5

SURVIVABILITY MEASURES

P_S FOR CONSTANT THREAT CHARACTERISTICS

PRIMARY CONTRIBUTOR IS ELUSIVENESS

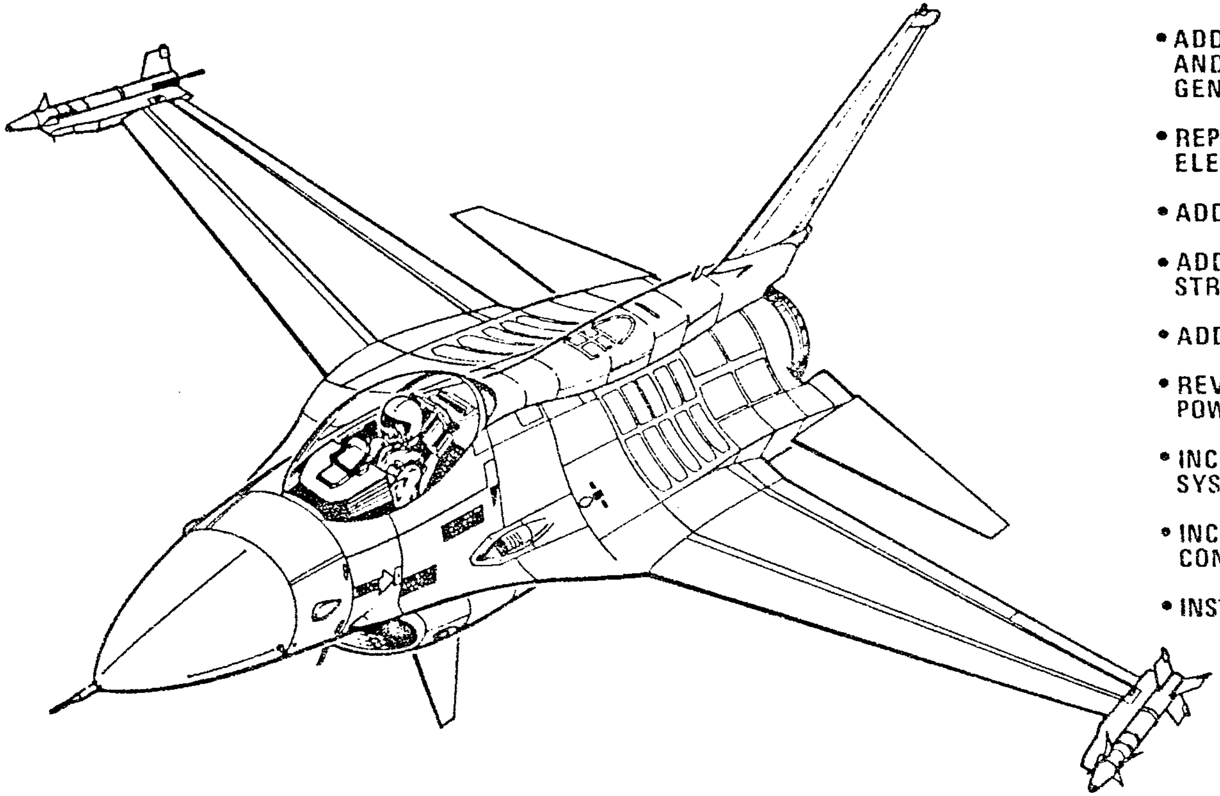
- o ALTITUDE PROFILE
- o EXPOSURE TIME
- o JINKING MANEUVERS
- o GROSS ACQUISITION TIME
- o STANDOFF RANGE
- o (FOR AIR-TO-AIR) LOSS OF ENERGY

SECONDARY CONTRIBUTOR IS FINE TRACKING ACCURACY

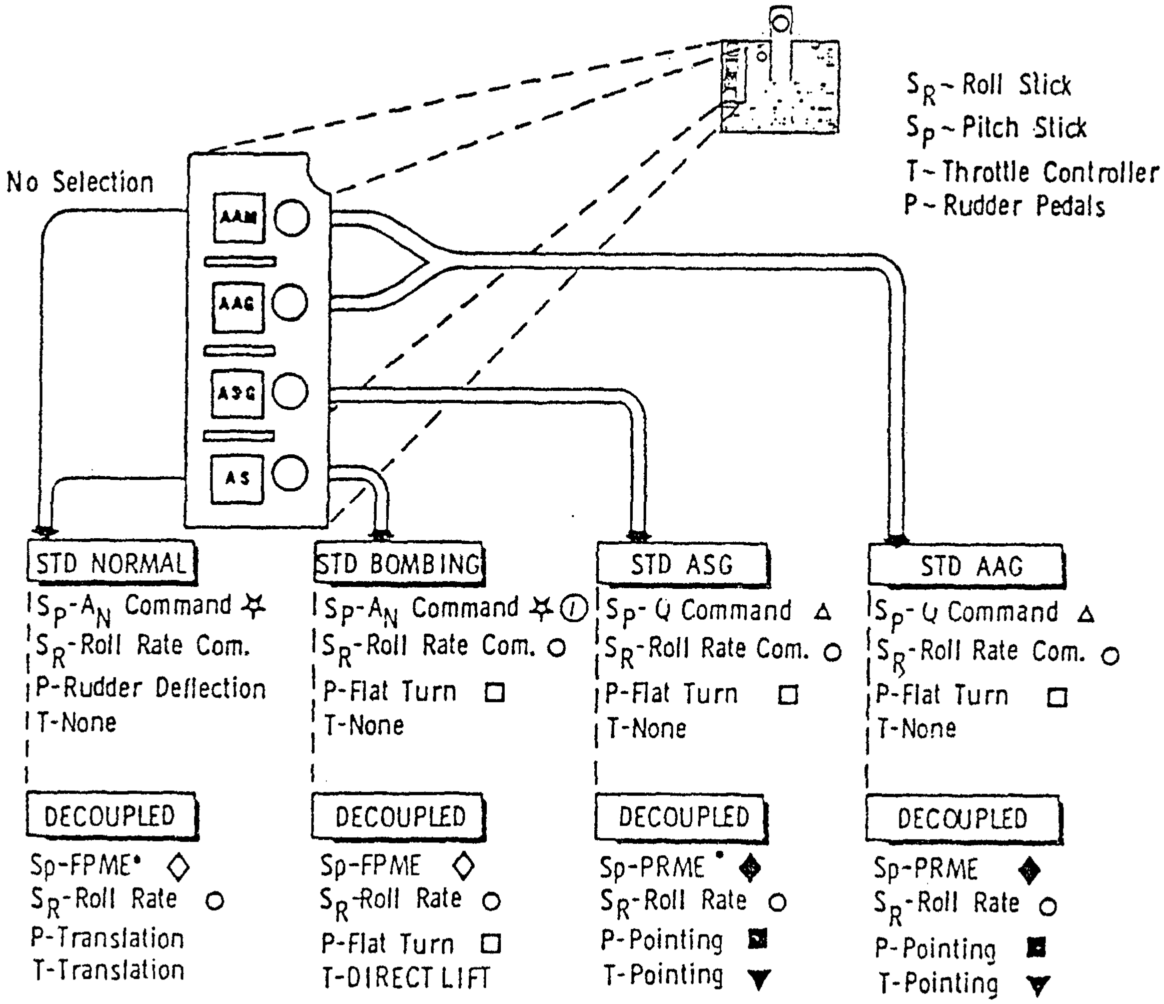
- o FINE TRACKING TIME
- o FLYING QUALITIES PARAMETERS

Figure 1

AFTI/F-16 Test Bed Modifications



- REPLACE FLIGHT CONTROL COMPUTER AND ACTUATOR INTERFACE UNIT
- REPLACE STORES MANAGEMENT CENTRAL INTERFACE UNIT
- ADD MULTIPURPOSE DISPLAYS AND PROGRAMMABLE DISPLAY GENERATORS
- REPLACE HUD AND HUD ELECTRONICS UNIT
- ADD NEW CONTROLLERS
- ADD CANARDS AND SUPPORT STRUCTURE
- ADD DORSAL FAIRING
- REVISE ELECTRICAL POWER SYSTEM
- INCREASE HYDRAULIC SYSTEM CAPACITY
- INCREASE ENVIRONMENTAL CONTROL SYSTEM CAPACITY
- INSTRUMENT AIRCRAFT



NOTE: Mode control axes with same symbols have identical control laws and gains

* FPME - FLIGHT PATH MANEUVER ENHANCEMENT
PRME - PITCH RATE MANEUVER ENHANCEMENT

① Identical outside bombing envelope only

Figure 2 AFTI/F-16 Multimode Flight Controller Commands

Figure 3 **AFTI/F-16 SCHEDULE**

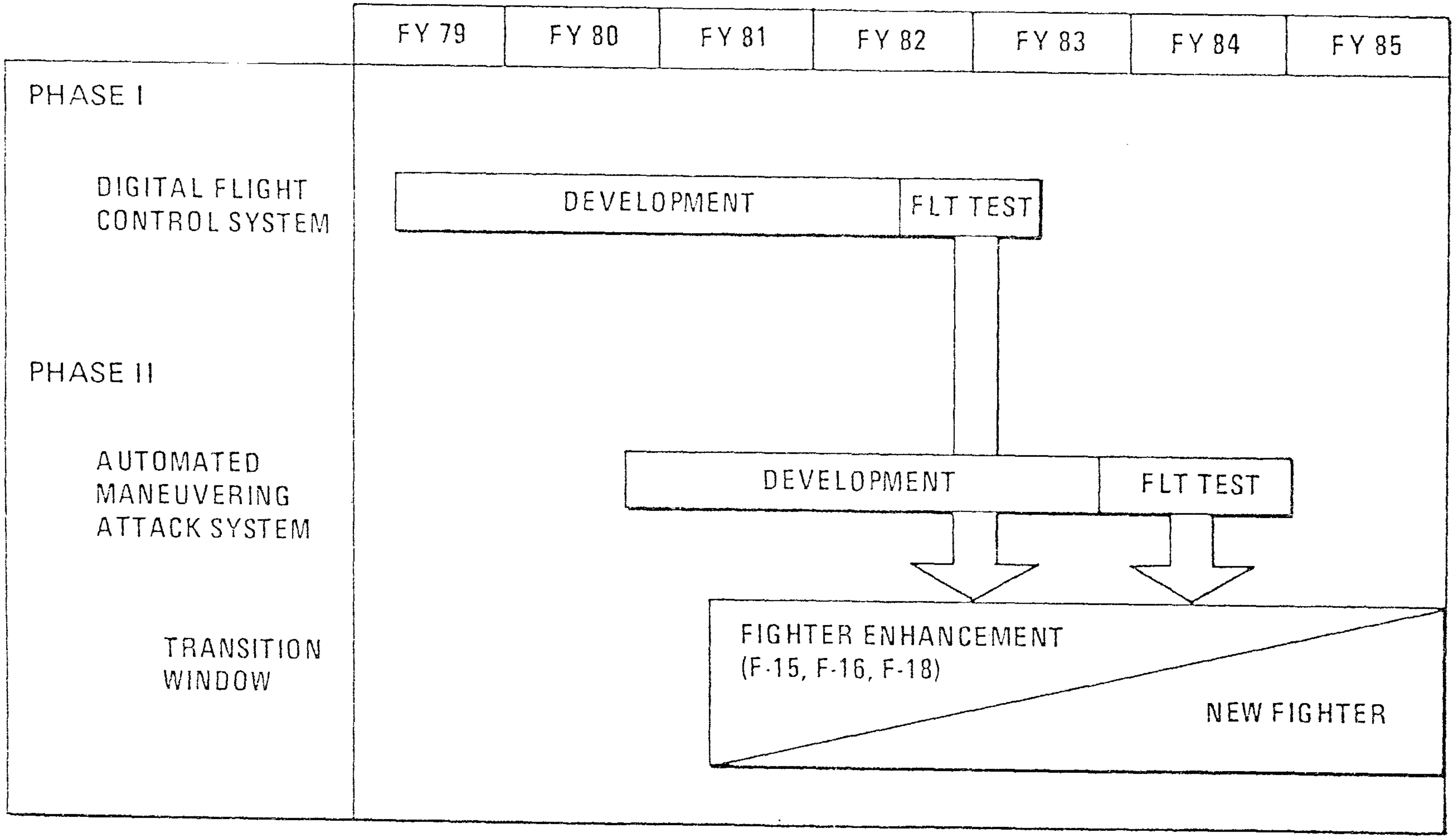


Figure 4

AUTOMATED MANEUVERING ATTACK

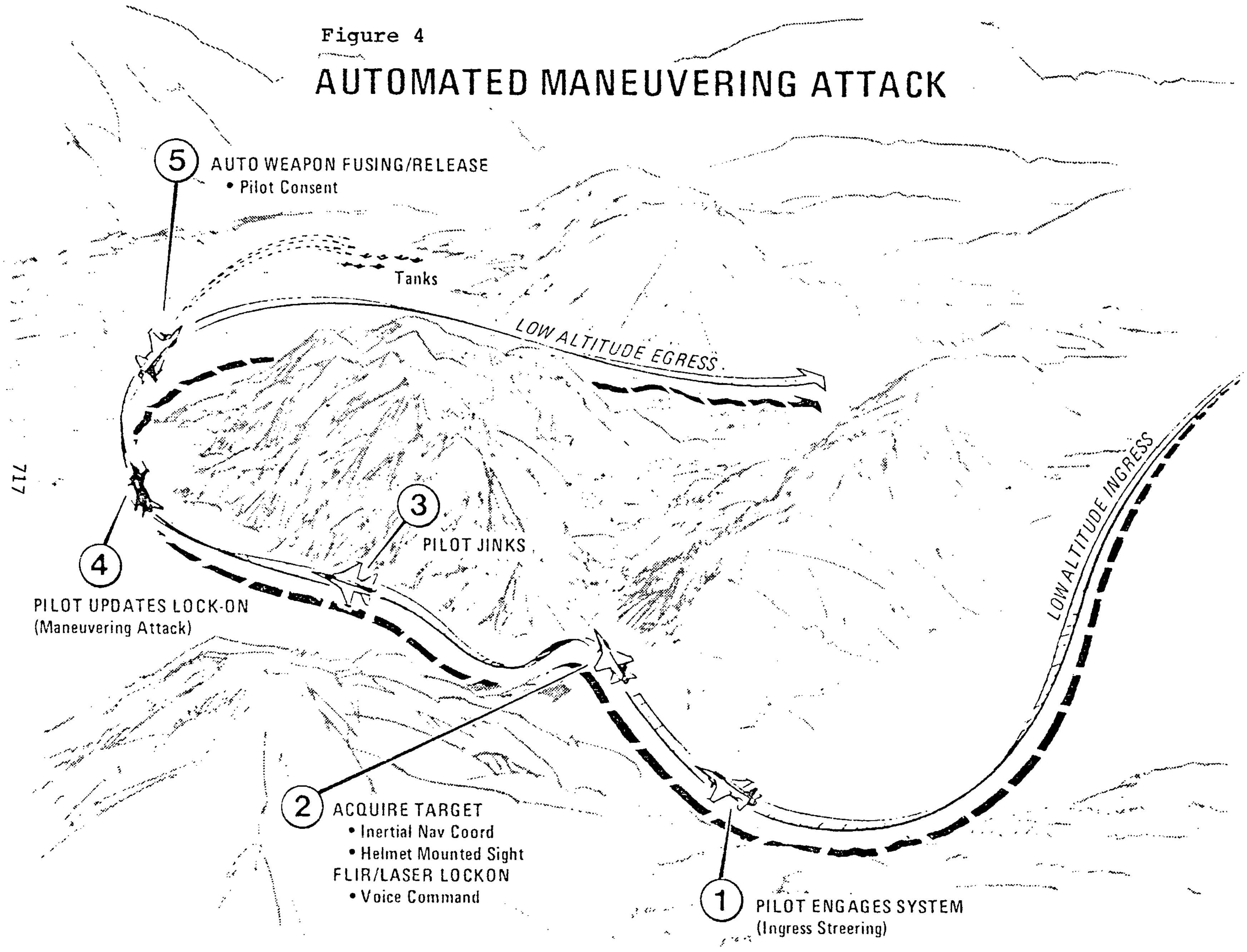
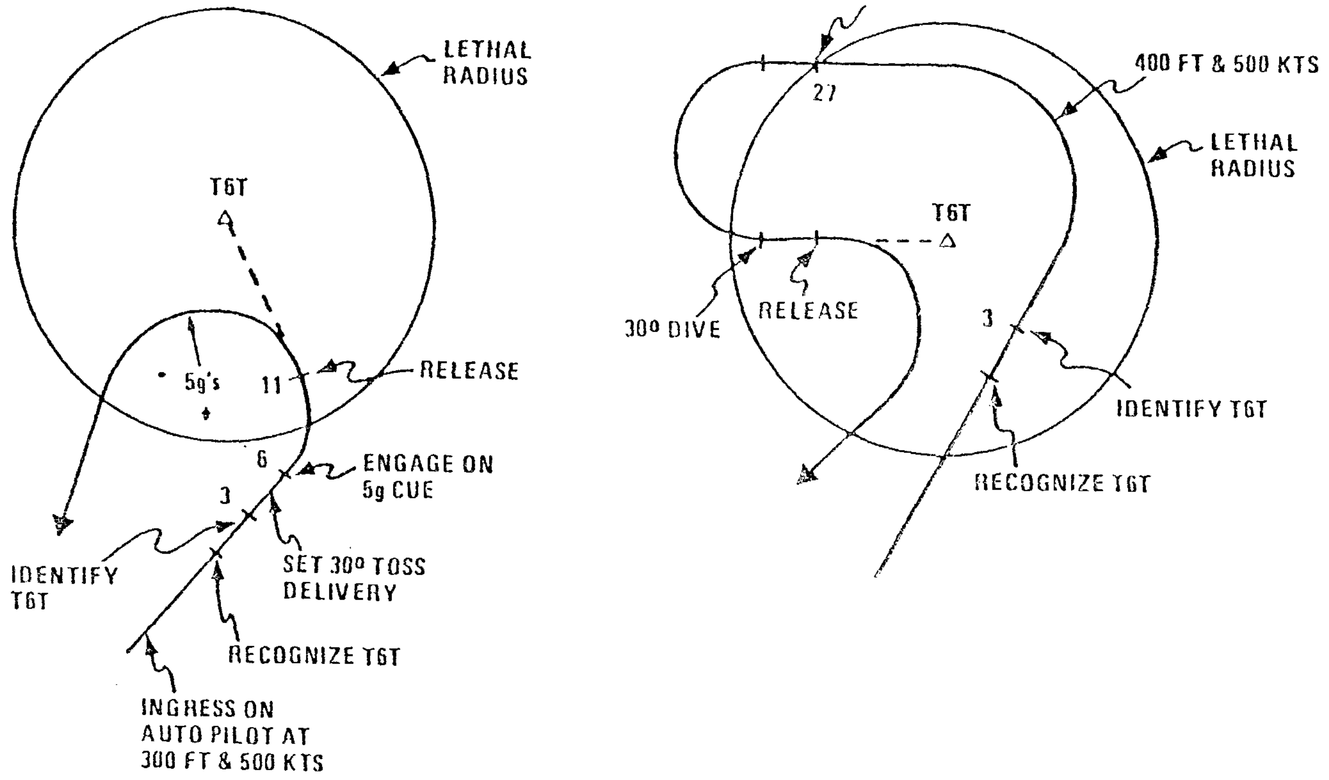


Figure 5 SURVIVABILITY COMPARISON PROFILES FOR AIR-TO-SURFACE BOMBING



AFTI/F-16
Automated Maneuvering Attack

F-16

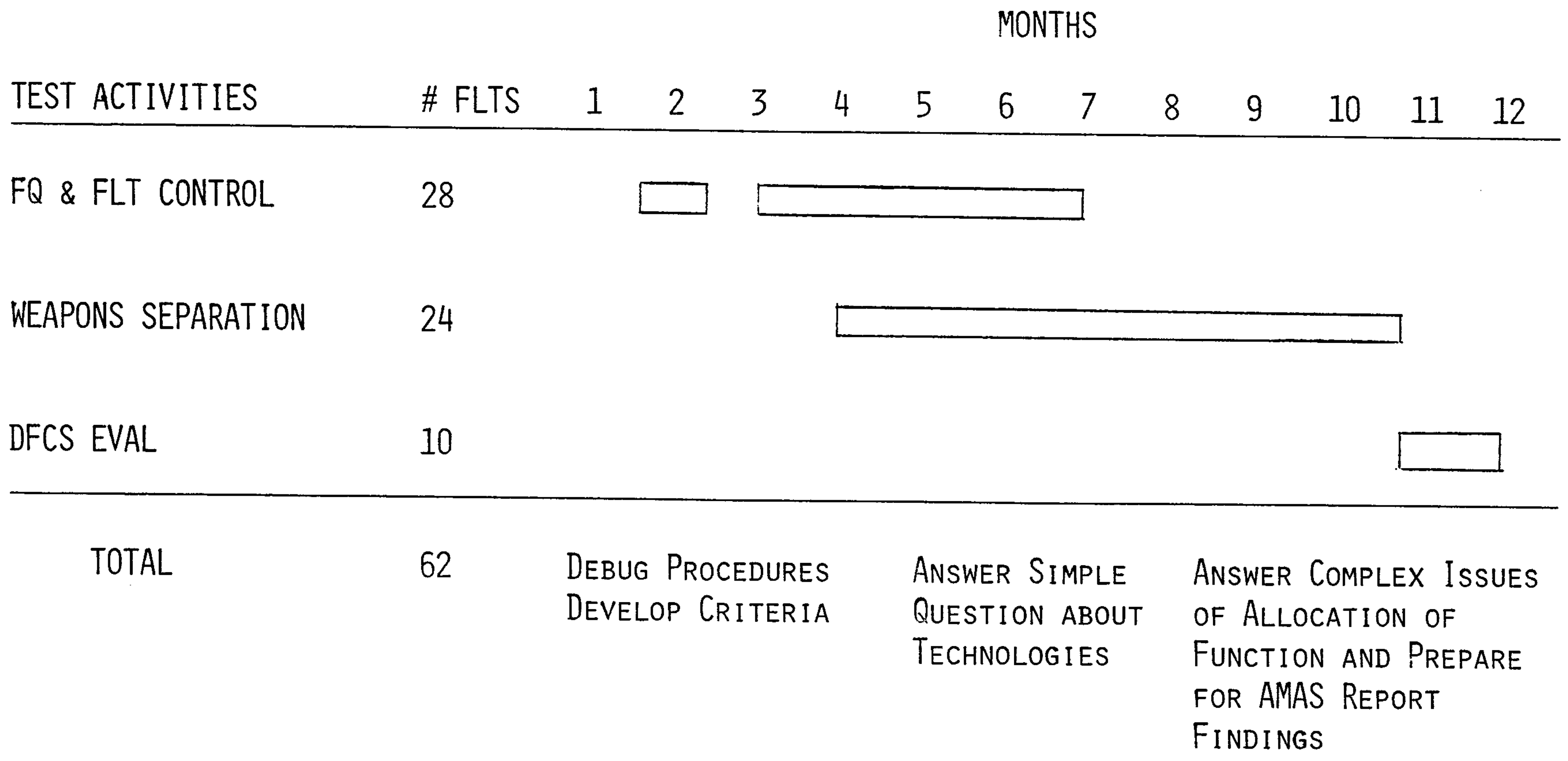


Figure 6 Workload Measurement Methods Development Schedule

APPENDIX A

CLOSED LOOP FLYING QUALITIES MANEUVERS

Formation

1. The evaluation of flying qualities during formation will be conducted in conjunction with other tests and by flying formation with chase aircraft while enroute to and from the test ranges.

2. Evaluate formation flying in the flight control modes specified in Table 8.2-2. Evaluate the flight control laws and gains implemented in each specified mode while flying a formation closed loop task.

3. Detailed procedures and conditions are not specified so as to encourage an unbiased evaluation of the utility of the various DFCS modes.

Aerial Refueling

1. Evaluate aircraft flying qualities during aerial refueling by utilizing standard aerial refueling procedures.

2. The flying qualities during refueling will be evaluated in the flight control modes specified. Evaluate the flight control laws and gains implemented in the specified flight control modes. Approach the close-in closed loop task in a conservative manner and determine the ability of the pilot to refuel utilizing conventional and decoupled control modes.

3. Detailed procedures are not specified so as to encourage an unbiased evaluation of the utility of the various DFCS modes.

Handling Qualities During Tracking (HQDT)

1. Confirm flight control mode, gun camera operation, F-stop and film magazine No. with test control.

2. Standby Reticle selected, depression angle set as directed by test control and HUD camera selected.

3. Tracking pilot should trim for level flight prior to the maneuver. Do not retrim during the maneuver. Target pilot should set up on speed and altitude and stabilize in a 30° bank.

4. Tracking pilot close to 1,500 ft. range.

5. Data "ON". In level flight accomplish a bias calibration for each film magazine by recording a short (1 sec) burst of film with the pipper on the aimpoint.

6. Tracking pilot call "Start Maneuver."

7. The target pilot should begin either a slow wind-up turn or a constant-turn at the test conditions as specified. For the wind-up turn, increase load factor at an approximate rate of 1 "g" per 10 second to reach about 5 "g's" in 40 seconds. Hold 5 "g's" for 5 seconds. Descend and use afterburner as necessary to maintain constant Mach number. For the constant-turn, the target should establish a constant altitude turn at the conditions specified. Use afterburner as necessary to maintain level flight and constant Mach number.

8. Tracking pilot squeezes the trigger to start the camera and briskly moves pipper on to target from 50 MIL reticle or farther.

9. Tracking pilot must persistently track the precision aimpoint agreed to prior to the flight, using the specified controllers. Attempt to hold 1000 to 2000 ft. range during the run.

10. Target pilot to call "Terminate" 45 seconds after starting maneuver.

11. At the call "Terminate", the tracking pilot should acknowledge and briskly move the pipper at least 50 MILS off the target and release the gun trigger.

12. Data "OFF."

NOTE: Either pilot may terminate the maneuver early if required by unusual or unsafe conditions. The tracking pilot is responsible for maintaining safe separation. During the maneuver the target pilot is responsible to maintain clearance from other aircraft.

13. Call gun camera film remaining to test control.

14. Change magazine if necessary and give magazine number to test control.

NOTE: Testing may indicate that some of the conditions in the run table are of more interest than others. Therefore, the run table should be viewed as a "menu" to select the specific test conditions. It is desirable to have 3 pilots fly 2 missions each of the HQDT maneuvers.

Air-to-Surface Tracking, Strafe

1. Select a reasonable strafe target (car, plane, strafe panel etc.) on the bombing range and set-up inbound at the conditions specified in the run table.

2. Perform a simulated strafe run by attempting to keep the CCIP pipper on the target during the run.

3. Use of decoupled options will be as specified in the preflight briefing.

4. The pilot shall simulate a one second burst when in range and execute a pull-up.

Air-to-Surface Tracking, Bomb

1. Set-up inbound to the target at 3500' AGL.

2. Upon reaching the point where the target is 10° below the horizon, pushover and track the target with the flight path marker.

3. Use only the controllers specified in the run table.

4. Recover from the dive at a safe altitude.

ILS

1. Conduct an ILS approach to landing utilizing standard ILS procedures.

2. Qualitatively evaluate the aircraft handling qualities during the ILS.

3. The pilot shall use his normal technique when performing this test. (i.e. HUD vs ADI, AOA as desired etc.)

Table 0.2-2

RUN NO.	AS/ MACH	ALT.	CONFIG.	EXT. LOAD	DFCS MODE	PRIORITY	MANEUVER		
SC-550	400kcas	-	CR	A/S	SASG	1	Air-to-Surface Gun Tracking		
SC-551	400kcas	-			DASG	1	Air-to-Surface Gun Tracking		
SC-552	460kcas	-			SASG	1	Air-to-Surface Gun Tracking		
SC-553	460kcas	-			DASG	1	Air-to-Surface Gun Tracking		
SC-554	520kcas	-			SASG	1	Air-to-Surface Gun Tracking		
SC-555	520kcas	-	CR	A/S	DASG	1	Air-to-Surface Gun Tracking		
SC-560	400kcas	3.5k	CR	A/S	SASB	1	Air-to-Surface Bomb Tracking		Stick Only
SC-561					SASB		Air-to-Surface Bomb Tracking		Stick & Pedals
SC-562					DASB		Air-to-Surface Bomb Tracking		Stick Only
SC-563					DASB		Air-to-Surface Bomb Tracking		Stick & Throttle
SC-564	400kcas				DASB		Air-to-Surface Bomb Tracking		Stick & Pedals
SC-565	500kcas				SASB		Air-to-Surface Bomb Tracking		Stick only
SC-566					SASB		Air-to-Surface Bomb Tracking		Stick & Pedals
SC-567					DASB		Air-to-Surface Bomb Tracking		Stick only
SC-568					DASB		Air-to-Surface Bomb Tracking		Stick & Throttle
SC-569	500kcas	3.5k	CR	A/S	DASB	1	Air-to-Surface Bomb Tracking		Stick & Pedals
SC-570	11°α	-	PA	A/A	SNRM	1	ILS & Landing		
SC-571	13°α	-	PA	A/A	SNRM	1	ILS & Landing		
SC-575	0.9	20k	CR	A/A	SAAG	1	HQDT - WUT		Stick only
SC-576					SAAG		HQDT - Const	3.4g	Stick only
SC-577					SAAG		HQDT - Const	3.4g	Stick & Pedals
SC-578					DAAG		HQDT-WUT		Stick only
SC-579					DAAG		HQDT - Const	3.4g	Stick only
SC-580					DAAG		HQDT - Const	3.4g	Stick & Throttle
SC-581	0.9				DAAG		HQDT - Const	3.4g	Stick Pedals
SC-582	0.6				SAAG		HQDT - WUT		Stick only
SC-583					SAAG		HQDT - Const	3.4g	Stick only
SC-584					SAAG		HQDT - Const	3.4g	Stick & Pedals
~~SC-585~~					DAAG		HQDT - WUT		Stick only
SC-586					DAAG		HQDT - Const	3.4g	Stick only
SC-587					DAAG		HQDT - Const	3.4g	Stick & Throttle
SC-588	0.6	20k			DAAG		HQDT - Const	3.4g	Stick & Pedals
SC-589	0.9	10k			SAAG		HQDT - WUT		Stick only
SC-590					SAAG		HQDT - Const	3.4g	Stick only
SC-591					SAAG		HQDT - Const	3.4g	Stick & Pedals
SC-592					DAAG		HQDT - WUT		Stick only
SC-593					DAAG		HQDT - Const	3.4g	Stick only
SC-594					DAAG		HQDT - Const	3.4g	Stick & Throttle
SC-595	0.9	10k	CR	A/A	DAAG	1	HQDT - Const	3.4g	Stick & Pedals
SC-530	250kias	20k	CR	A/A	SNRM	1	Refueling		
SC-531	275	20k			SNRM		Refueling		
SC-532	300	20k			SNRM		Refueling		
SC-533	250	30k			SNRM		Refueling		
SC-534	275	30k			SNRM		Refueling		
SC-535	300kias	30k			SNRM		Refueling		
SC-536	Opt.	Opt.		A/A	DNRM		Refueling		
SC-537	Opt.	Opt.		A/S	SNRM		Refueling		
SC-538	Opt.	Opt.	CR	A/S	DNRM	1	Refueling		
SC-540	-	-	CR	A/A	SNRM	1	Formation		
SC-541	-	-	CR	A/S	SNRM	1	Formation		
SC-542	-	-	CR	A/A	DNRM	1	Formation		
SC-543	-	-	CR	A/S	DNRM	1	Formation		

APPENDIX B

8.7 WEAPON SEPARATION MANEUVERS

8.7.1 Test Objectives

To meet the design goal of improved weapon delivery accuracy combined with increased survivability, the AFTI/F-16 program will require that weapons be released from the aircraft in a variety of attitudes and accelerations not common to the F-16. It is the objective of the weapons separation testing to provide store separation clearance for the most critical flight/release conditions anticipated for the AFTI/F-16 program. This will be accomplished by ejecting BDU-33 D/B practice bombs from SUU-20 dispensers at specified conditions.

8.7.1.1 Test Approach. Data reduction by the photo-imaging techniques used in the F-16 FSD program (photo data analysis system) will be required to provide store separation clearance for the AFTI/F-16 aircraft. Correlations of the flight test data and the predicted separation characteristics will be made after each flight before moving to a more critical separation condition. The flight test conditions will be adjusted if necessary to assure adequate aircraft safety. These steps will be continued until satisfactory store separation has been verified at the most critical flight and release condition anticipated for the AFTI/F-16 program. Weapon separation clearance drops will be made at higher lateral accelerations and angles of sideslip than listed in Table 8.7-1 if Weapon Effectiveness objectives indicate a payoff and if clearance approval for the higher levels is obtained from AFATL/DLJC. The higher levels to be considered as a goal are ±2.0g sideforce and ±8.0 degrees sideslip.

In most cases one subcritical release will be required before a critical release is executed. In the case of canard deflection, however, the level of uncertainty is much higher with respect to flow field changes at the SUU-20. For that reason, up to three subcritical separations may be required before each critical drop.

The estimated data turnaround time between subcritical and critical separation flights is now projected to be approximately 2-1/2 to 3 weeks. This period will allow for film processing (AFFTC Edwards AFB), data extraction (PDAS, Pt. Mugu) and trajectory analyses (GD/FW). In some cases a simple review of inflight drop films will suffice for clearance of subsequent flights. In most cases the data can be pipelined; i.e., several data sets in review/analysis process simultaneously so that separation flights evaluating different effects can be inter-leaved or combined without delaying flight scheduling. However, in planning for significantly critical drops, the full 2-1/2 to 3 weeks interval should be allowed.

Following any store separation build-up test, the contractor (with Air Force concurrence) will issue a clearance to proceed to the next planned test in a test series after data analyses of onboard film and/or magnetic tape have been performed.

8.7.3 Test Conditions

The test conditions for the weapon separation clearance program are contained in Table 8.7-1. Each test point requires that all six bombs from the SUU-20 be released in ripple fashion. This will be accomplished by selecting the "RIPPLE" mode of SUU-20 operation. Runs that call for asymmetric release conditions, (Bank angle $\neq 0°$ or $180°$, $N_y \neq 0$, or $\beta \neq 0$) shall require that bombs be dropped from both wings simultaneously.

8.7.2 Test Maneuvers and Procedures

Store separation flight testing is not usually thought of as a discipline requiring the use of classical flight test maneuvers. This is because data prior to release is relatively unimportant compared to data at and immediately after release which defines the aircraft and store separation characteristics.

However, AFTI/F-16 store separation tests at specified Mach no., altitude, normal acceleration, dive angle, bank angle, canard toe-in angle and lateral acceleration not only require precise flying but a high degree of timing so that the release and impact occur on range. (Radar vectoring and space positioning will be utilized to ensure proper range safety).

Several separation tests require releases in a 60 degree dive angle at 0.9 Mach and at the highest practical $\underline{q}$. Altitude loss during dive recovery (5500 feet) was obtained by assuming throttle to idle after release, speed brakes fully extended (no canard toe-in) and a 4 "g" recovery (nominal S.L. temperature - 25°C). To this value, 1,500 feet was added to account for ripple release time, pilot reaction time and time to roll the aircraft to a satisfactory recovery bank angle. Another 1500 feet was added to provide required minimum ground clearance. Thus an initial weapon release altitude of 8,500 feet AGL is obtained. This altitude was also used as the release altitude for related build-up separation at 30 degree dive angles.

A. Level Flight Release

1. Stabilize at the flight conditions specified and select specified flight control mode.

2. When specified, extend speedbrakes to obtain the proper canard deflection. Use power as required.

3. When on conditions release one SUU-20 load of BDU-33s.

B. Direct Lift Command/Level Flight Release

1. Stabilize at a preselected Mach and altitude above release altitude. Select specified flight control mode.

2. Conduct a constant Mach (approximately -20°) dive. Pull out from the dive with full nose up direct lift command such that the aircraft is level at approximately 10k.

3. Maintain full nose up direct lift command and command bomb release as the nose passes through the horizon.

C. Release in a Stabilized Dive

1. Select the specified flight control mode and extend the speedbrakes to the specified canard toe-in angle (if specified) at a preselected altitude.

2. Roll into a dive at the preselected altitude and Mach number.

3. Quickly establish dive angle and make power adjustments to control Mach number such that release may be commanded at specified dive angle Mach number and altitude.

D. Non-Wings Level Release

1. Stabilize briefly at the specified flight conditions, select specified flight control mode and extend speed brake to full toe-in angle. Use power as required.

2. Pull up into a climb, roll to the specified bank angle, pull to +1.5 "g" (Nz) and allow the nose to drift down through the horizon.

3. When bank angle and Nz are on conditions and pitch angle is between 0° and -30°, command release.

4. Maintain flight conditions for 3 seconds, roll and increase "g" to vector aircraft away from bomb trajectory.

E. Level Flight/Flat Turn Release

1. Stabilize at the specified flight conditions and select specified flight control mode.

2. Apply left rudder pedal slowly and stabilize at the specified lateral acceleration (Ay). Maintain wings level.

3. When on conditions, command the release.

F. Stabilized Dive/Flat Turn Release

1. Select the specified flight control mode and briefly stabilize at a preselected altitude.

2. Roll into a dive at the preselected altitude and Mach number.

3. Quickly establish dive angle and make power adjustments to control Mach number. Apply left rudder pedal and stabilize at the specified lateral acceleration (Ay).

4. Command release at specified dive angle, Mach number, Ay and altitude.

G. Non-Wings Level/Flat Turn Release

1. Stabilize briefly at the specified flight conditions and select specified flight control mode.

2. Pull up into a climb, roll left to the specified bank angle, pull to 1.5 "g" (Nz), apply left rudder pedal and stabilize at the specified lateral acceleration (Ay).

3. Allow the nose to drift down through the horizon.

4. When bank angle, Nz and Ay are on conditions and pitch angle is between 0° and -30°, command release.

5. Maintain flight conditions for 3 sec., release rudder pedal, roll and increase "g" to vector aircraft away from bomb trajectory.

H. Direct Lift Command/Flat Turn/Level Flight Release

1. Stabilize briefly at a preselected Mach and altitude above release altitude. Select the specified flight control mode.

2. Conduct a constant Mach (approximately -20°) dive. Pull out of the dive with full nose up direct lift command such that the aircraft is level at approximately 10k. Prior to attaining level flight apply left pedal command to attain specified lateral acceleration (Ay).

3. As the nose passes through the horizon with full nose up direct lift command and at specified Ay, command bomb release.

I. Stabilized Dive/Direct Lift/Flat Turn Release

1. Stabilize briefly at a preselected Mach and altitude and select specified flight control mode.
2. Conduct a 60° dive such that the specified flight conditions and flight control system inputs occur at approximately 8.5k AGL.
3. Prior to reaching 8.5k command full nose-up direct lift command and left pedal command to 0.5 "g" Ay.
4. At approximately 8.5k AGL with full nose-up direct lift command and 0.5 "g" Ay, command bomb release.
5. Release L/H controller and pedal inputs and recover from dive.

8.7.2.1 Maneuver Tolerances. Selection of parameter limits in flight test is always difficult since too stringent tolerances can make the maneuver almost impossible for the pilot to fly and too wide a tolerance can result in the collection of invalid data. In either event a costly repeat of a run can result.

The tolerances set forth below are proposed as a goal so as to encourage the performance of precise tests with resulting valid data. Only the magnitude of each tolerance is shown; the sign convention does not refer to direction.

Bank angle	±5°	(±15° non wings level)
Normal Load Factor	±0.2"g"	(±0.5 at "g's" > 2.0)
Lateral Load Factor	+0.2, -0g	
Sideslip Angle	±1°	
Canard Toe-in Angle	±1°	
Roll Rate	0°/sec	
Pitch Angle	±5°	
Mach Number	±.03	
Altitude	±1000 feet	

8.7.2.2 Maneuver Limits. The following maneuver limits are set up as a guide for clearing the AFTI/F-16 for BDU-33 weapon releases. Upon completion of Table 8.7-1, the AFTI/F-16 will be cleared to release BDU-33s within the following limits:

	Wings Level	Non-Wings Level
Bank angle	0°	±180°
Normal Load Factor	0.5 - 5.0 "g"	1.5 - 5.0"g"
Lateral Acceleration	±0.5"g"	±0.5"g"
Angle of Sideslip	±1.0°	±1.0°
Canard Toe-in angle	0-27°	0-27°
Roll Rate	0°/sec	0°/sec
Angle of Attack	+10°	+10°

Table 8.7-1 WEAPON RELEASE CONDITIONS

RUN NO.	MACH	ALT.	PRIORITY	DFCS MODE	Nz	PITCH ANGLE	BANK ANGLE	CANARD TOE-IN	Ny	δ_{Vc}	δ_{TEF}	SUU-20 LOADS	PROCEDURE	AOA
WS-01	0.4	10K	1	DASB	1.3	0°	0°	0°	0	0	20	1	8.7.2 B	4.7
02	0.8	5K		SASB	1.0	0°	0°	0°	0	0	-2	1	8.7.2 A	1.5
03	0.9	5K		DASB	1.0	0°	0°	0°	0	0	-2	1	8.7.2 A	1.3
04	0.9	8.5K AGL		SASB	0.5	-60°	0°	0°	0	0	-2	1	8.7.2 C	1.0
WS-10	0.8	5K		SASB	1.0	0°	0°	9°	0	0	-2	1	8.7.2 A	1.6
11	0.8	5K		DASB	1.0	0°	0°	18°	0	0	-2	1	8.7.2 A	1.8
12	0.8	5K		SASB	1.0	0°	0°	27°	0	0	-2	1	8.7.2 A	1.9
13	0.8	8.5K AGL		DASB	0.8	-30°	0°	27°	0	0	-2	1	8.7.2 C	1.8
14	0.8	8.5K AGL		SASB	0.5	-60°	0°	27°	0	0	-2	1	8.7.2 C	1.5
WS-20	0.9	5K		SASB	1.0	0°	0°	14°	0	0	-2	1	8.7.2 A	1.5
21	0.9	5K		DASB	1.0	0°	0°	27°	0	0	-2	1	8.7.2 A	1.7
22	0.9	8.5K AGL		SASB	0.8	-30°	0°	27°	0	0	-2	1	8.7.2 C	1.6
23	0.9	8.5K AGL		DASB	0.5	-60°	0°	27°	0	0	-2	1	8.7.2 C	1.3
WS-30	0.9	8.5K AGL		SASB	1.5	0°to-30°	60°	27°	0	0	-2	2	8.7.2 D	2.2
31	0.9	8.5K AGL		DASB	1.5	0°to-30°	120°	27°	0	0	-2	2	8.7.2 D	2.2
32	0.9	8.5K AGL		SASB	1.5	0°to-30°	180°	27°	0	0	-2	1	8.7.2 D	2.2
WS-40	0.8	5K		SASB	1.0	0°	0°	0°	0.25	5.8	-2	2	8.7.2 E	1.5
41	0.8	5K		DASB	1.0	0°	0°	0°	0.5	11.6	-2	2	8.7.2 E	1.5
42	0.8	8.5K AGL		SASB	0.5	-60°	0°	0°	0.5	12.6	-2	2	8.7.2 E	1.0
43	0.9	5K		DASB	1.0	0°	0°	0°	0.25	4.5	-2	2	8.7.2 E	1.3
44	0.9	5K		SASB	1.0	0°	0°	0°	0.5	9.0	-2	2	8.7.2 E	1.3
45	0.9	8.5K AGL		DASB	0.5	-60°	0°	0°	0.5	9.9	-2	2	8.7.2 F	1.0
WS-50	0.9	8.5K AGL		SASB	1.5	0°to-30°	60°	0°	0.5	10.7	-2	2	8.7.2 G	1.8
51	0.9	8.5K AGL		DASB	1.5	0°to-30°	120°	0°	0.5	10.7	-2	2	8.7.2 G	1.8
52	0.9	8.5K AGL		SASB	1.5	0°to-30°	180°	0°	0.5	10.7	-2	2	8.7.2 G	1.8
53	0.9	8.5K AGL		DASB	2.8	0°	0°	0°	0.5	8.6	20	2	8.7.2 H	1.4
54	0.9	8.5K AGL	1	DASB	2.4	-60°	0°	0°	0.5	8.7	20	2	8.7.2 I	1.0

APPENDIX C

RESPONSE TO TURBULENCE MANEUVER

One of the PVI evaluation maneuvers listed in Table 1 is the response to turbulence maneuver. This maneuver will be used to evaluate the gust alleviation characteristics of five of the flight control modes at two airspeeds each. Gust alleviation differs in each mode in the pitch axis only. The gun modes resist pitch rate upsets, the bombing modes resist flight path disturbances, and the normal modes provide improved ride qualities at the pilot station.

The maneuver is approximately three minutes long to collect enough data to perform a spectral analysis. The piloting task is to keep the wing level (bank angle regulation) while minimizing longitudinal control inputs. This may allow an analysis of pilot dynamics for the bank angle regulation task for ten different runs.

The vertical gust velocity (disturbance) time history will be calculated from angle of attack and normal acceleration or INS vertical velocity. The power spectral density of vertical gust velocity and responses of normal acceleration and pitch rate will then be computed and combined to produce frequency response functions. In the lateral axis the pilot describing function of lateral stick force to bank angle error may also be computed.

The flight conditions will all be on the same flgiht over the same terrain at one thousand feet above ground level for each Mach number. Test conditions are detailed in the following table.

RUN NO.	AS/ MACH	ALT	CONFIG	EXT. LOAD	DFCS MODE	PRIORTY
GR-01	400KCAS	1KAGL	CR	A/A	SNRM	1
GR-02	400KCAS	1KAGL	CR	A/A	SAAG	1
GR-03	400KCAS	1KAGL	CR	A/A	SASB	1
GR-04	400KCAS	1KAGL	CR	A/A	DAAG	1
GR-05	400KCAS	1KAGL	CR	A/A	DASB	1
GR-06	520KCAS	1KAGL	CR	A/A	SNRM	1
GR-07	520KCAS	1KAGL	CR	A/A	SAAG	1
GR-08	520KCAS	1KAGL	CR	A/A	SASB	1
GR-09	520KCAS	1KGAL	CR	A/A	DAAG	1
GR-10	520KCAS	1KGAL	CR	A/A	DASB	1

APPENDIX D

9. DFCS EVALUATION PHASE

The DFCS evaluation test objectives are to (1) conduct an evaluation of the operational usefulness of manual control of DFCS task tailored modes and (2) conduct an evaluation of the capability of the aircraft to meet the air-to-air and air-to-surface mission performance requirements. These objectives will be accomplished on dedicated evaluation flights and/or as secondary objectives on DFCS development flights (i.e., remaining portions of weapon separation flights). Priority will be given to A/S bombing.

The basic approach to evaluation testing will be as defined in this section. The scheduling (when to test) and test scope (what to test) will be determined by the AFTI/F-16 Joint Test Force (JTF), with AFTI/F-16 ADPO concurrence, based on the following factors: (1) level of maturity of the specific DFCS modes, (2) results of DFCS development tests, (3) results of pilot experience during DFCS development tests and (4) flight time available. In general DFCS evaluation tests will have a lower priority than DFCS development tests required to clear the flight envelope for the AMAS flight test phase.

9.1 DFCS EVALUATION AND OPERATIONS DEVELOPMENT

9.1.1 Test Objectives/Approach

The objective of the tests contained in this section is to determine the operational usefulness and evaluate the mission capabilities of the various AFTI/F-16 technologies. This will be accomplished by flying simulated/actual air-to-air and air-to-surface weapon delivery runs, emphasizing use of the AFTI - unique flight control modes and PVI advancements. In that the Integrated Fire and Flight Control (IFFC) system will not be installed until the AMAS phase of testing, only manually flown profiles are included.

The emphasis of the DFCS Evaluation and Operations Development tests is on "pipper effectiveness"; that is, the relative degree of ease or difficulty in using the aircraft controls to position the aircraft, and in turn, the pipper. Runs are defined using both conventional maneuvering techniques and decoupled control mode options. Successful accomplishment of the tests outlined herein should yield pilot recommendations with supporting quantitative data on relative worth of the various DFCS modes and PVI improvements as they relate to manual weapon deliveries.

9.1.2 Test Maneuvers and Procedures

The runs/maneuvers to be used during the DFCS Evaluation/Operations Development tests are divided into three areas: (1) air-to-surface bombing, (2) air-to-surface gunnery, and (3) air-to-air gunnery. Not all of the runs call for use of decoupled modes to maneuver the aircraft to the point of weapons release. Some runs are designed to evaluate potential improvements using only conventional maneuver techniques that may be realized with the various DFCS modes. All runs described herein require a fully operational avionics system. Details on the switchology and avionics operating procedures required to use each weapon delivery mode will be included in the daily run cards. Additional information on these modes can be found in the F-16 Avionic System Manual (Block 10).

Of least importance in the accomplishment of the runs in this section is the need to precisely meet the dive angles, airspeeds, altitudes, etc., specified in the tables. These numbers are included only so as to define a range for parameters that may have a bearing on the effectiveness of a given mode to perform a specific task. In cases where it becomes apparent that varying airspeed, dive angle etc., will have no effect on the results, the remaining runs should be deleted. Conversely, if it appears that particular modes would be useful at conditions other than those specified in the run table, the run tables will be revised.

9.1.2.1 Air-to-Surface Bombing

1. Flat Turn Effectiveness - The aircraft shall be vectored inbound to the target along a ground track that will position it at the range and target angle-off specified in the run table. Upon reaching this point, the pilot should be stabilized on conditions (speed, heading, and altitude). Test control will call "maneuver" and the pilot will immediately use the specified technique to turn the aircraft and place the CCIP bomb fall line (BFL) over the target. The pilot should continue to "track" the target with the BFL and fly the run through to simulated/actual weapon release. The maneuver techniques to be used are as follows:

    a. Flat Turn: Full input flat turn (coupled or decoupled) to align with the target with pitch and roll stick inputs sufficient to maintain a wings level, constant altitude maneuver.

    b. Conventional: Target alignment accomplished strictly using pitch and roll stick inputs.

    c. Combined: Combined use of flat turn and conventional techniques in a fashion determined by the pilot to most rapidly line-up on the target. (Example, roll and pull for gross acquisition, flat turn for fine tracking inputs.)

The primary goal of this series of runs is to determine the most effective maneuver technique for target alignment as a function of heading change required.

Table 9.1-1 FLAT TURN EFFECTIVENESS

RUN NO.	TARGET RANGE	BEARING ANGLE OFF	MANEUVER TECHNIQUE	ACCOMPLISHED FLIGHT NO.
EV001	TBD	TBD	Flat Turn	
002			Conventional	
003			Combined	
004			Flat Turn	
005			Conventional	
006			Combined	
007			Flat Turn	
008			Conventional	
009			Combined	
010			Flat Turn	
011			Conventional	
012			Combined	
013			Flat Turn	
014			Conventional	
015			Combined	

All runs shall be performed at 480 knots/2KAGL, in either SASB or DASB. The option of using coupled or decoupled flat turn is left to the pilot's discretion. C-band tracking and HUD film is required.

2. Pop-Up Attack Evaluation - Set up inbound to the target sufficiently offset to perform a pop-up attack. Ingress altitude shall be between 200' and 2000' AGL. (Based on pilot proficiency and specific run objectives.) At the point the target is 30° angle off, initiate the pop-up. Upon rolling out on final, use flat turn inputs to line-up the BFL with the target. Follow the run through to simulated/actual release. Evaluate the utility of the decoupled flat turn mode in the fine tracking phase of this type of profile. Repeat the runs in DASB and evaluate usefulness of the direct lift function on the left hand controller during the pop-up.

Table 9.1-2 POP-UP ATTACK EVALUATION

RUN NO.	KCAS	AGL ALT	DFCS MODE	FCC MODE	REMARKS	ACCOMPLISHED FLIGHT NO.
EV020	480	A/R	SASB	CCIP		
021	550	│	│	CCIP		
022	480	│	│	DTOS		
023	550	│	SASB	DTOS		
024	480	│	DASB	CCIP	Evaluate Direct Lift	
025	550	│	│	CCIP	function on left	
026	480	│	│	DTOS	hand controller	
027	550	A/R	DASB	DTOS	during pop-up	

The use of coupled or decoupled flat turn is per plot discretion. HUD film and C-band tracking is required.

3. DTOS (Dive Toss) Evaluation - Set up 3000 to 5000 feet above the specified designate altitude at an angle off such that a roll-in to the target will yield a dive angle reasonably near that specified in the run table. Roll-in on the target, and once established on final in the dive, use the decoupled options of flat turn and direct lift to place the flight path marker (FPM), target designator (TD) box combination over the target. Designate with the weapon release switch and begin a 4g wings level pull. After designation, null out any steering error using flat turn. If desired, repeat the run in SNRM using standard maneuver techniques as a baseline comparison. Also vary the use of snow plow canards from run to run as specified in the run table. Evaluate the suitability of the AFTI/F-16 controls and displays to this type of delivery.

Table 9.1-3 DTOS EVALUATION

RUN NO.	MAX DIVE SPEED KCAS	DRAG MODE	DFCS MODE	FCC MODE	DIVE ANGLE	DESIGNATE ALT (AGL)	ACCOMPLISHED FLIGHT NO.
EV030	400	CONV	DASB	DTOS	30°	5K	
031	400	MODUL					
032	480	CONV					
033	480	MODUL					
034	550	CONV					
035	550	MODUL			30°	5K	
036	400	CONV			45°	7.5K	
037	400	MODUL					
038	480	CONV					
039	480	MODUL					
040	550	CONV					
041	550	MODUL			45°	7.5K	
042	480	CONV			60°	10K	
043	480	MODUL					
044	550	CONV					
045	550	MODUL	DASB		60°	10K	

The use of coupled or decoupled flat turn is per pilot discretion. HUD film and C-Band tracking is required. If speedbrake effectiveness is insufficient to hold down speed, pilot shall use full speedbrakes, idle power, and accept terminal dive airspeed.

4. CCRP (Continuously Computed Release Point) Evaluation - This run is designed to simulate a blind delivery using the radar for target designation. The pilot shall attempt to maintain "eyes in the cockpit" so as to best evaluate the overall layout and suitability of the cockpit and controls to this type of delivery technique. Set-up inbound to the target at approximately 15 to 20 miles. Upon turning final, select the appropriate steerpoint, delivery mode etc., and continue inbound to the target. Refine radar cursor placement (target designation) as required throughout the run, and use flat turn inputs to correct steering errors. Evaluate the use of direct lift inputs to maintain altitude. Fly the run through to simulated/actual release.

Table 9.1-4 CCRP EVALUATION

RUN NO.	KCAS	AGL ALT	DFCS MODE	FCC MODE	ACCOMPLISHED FLIGHT NO.
EV050	400	3k	DASB	CCRP*	
EV051	480	|	|	|	
EV052	550	3k	DASB	CCRP	

The use of coupled or decoupled flat turn shall be per pilot descretion. HUD film, MPD video, and C-Bank tracking is required.

*An offset aimpoint shall be used for this run with the pilot FCNP entries made after turning final.

5. LADD (Low Altitude Drogue Delivery) Evaluation - This maneuver is included so as to evaluate the pilots ability to follow the vertical and directional steering cues of the LADD mode using pitch stick left hand controller (direct lift) and rudder pedal (flat turn) inputs. Before flight, enter a pull-up range of 10,000 feet, and zero quantity MK-106 weapons. Inflight, set-up 15 miles inbound to the target and refine cursor placement (target designation) in the expand radar mode. Use flat turn inputs to null out any steering errors. Monitor the HUD symbology for the pull-up anticipation cue. On command from the HUD, perform a 4g pull to a 45° climb angle. Null out any vertical steering errors using pitch stick and left hand controller inputs, but use only flat turn inputs to correct directional errors. Fly the run through to simulated/actual release. Evaluate the usefulness of the AFTI/F-16 peculiar controls/displays in this type of weapon delivery profile.

Table 9.1-5 LADD EVALUATION

RUN NO.	KCAS	AGL ALT	DFCS MODE	ACCOMPLISHED FLIGHT NO.
EV055	480	200'-2000'	DASB	
EV056	550	200'-2000'	DASB	

The use of coupled or decoupled flat turn shall be per pilot discretion. HUD film is required.

9.1.2.2 Air-to-Surface Gunnery

1. Flat Turn - Fine Tracking - Set up inbound to the target so as to meet the conditions specified in the run table. Perform strafing pass on a standard 20' x 20' target on the Edwards gunnery range. Use decoupled flat turn inputs to correct any directional pipper errors. The pilot shall evaluate the wide field-of-view (FOV) Head Up Display (HUD) and the utility of decoupled flat turn to make steering/pipper corrections.

Table 9.1-6 FLAT TURN FINE TRACKING

RUN NO.	KCAS	DIVE ANGLE	DFCS MODE	FCC MODE	ACCOMPLISHED FLIGHT NO.
EV060	400	10°	SASG	STRF	
061	460				
062	520				
063	400	15°			
064	460				
065	520				
066	400	20°			
067	460				
068	520	20°	SASG	STRF	

The use of coupled or decoupled flat turn shall be per pilot discretion. HUD film is required.

2. Strafe Pitch/Yaw Pointing - Set up inbound to the target at the conditions in the run table, with no pitch or yaw pointing inputs. The pilot should intentionally place the aircraft velocity vector somewhat off the intended target. While proceding inbound, and especially once in-range, pitch and yaw pointing commands should be used to place the pipper on the target. Evaluate the feasibility of using pitch and yaw pointing inputs as a control device for strafe pipper positioning.

Table 9.1-7 STRAFE PITCH/YAW POINTING

RUN NO.	KCAS	DIVE ANGLE	DFCS MODE	FCC MODE	ACCOMPLISHED FLIGHT NO.
EV070	400	10°	DASG	STRF	
071	460	│	│	│	
072	520	10°	│	│	
073	400	15°	│	│	
074	460	│	│	│	
075	520	15°	│	│	
076	400	20°	│	│	
077	460	│	│	│	
078	520	20°	DASG	STRF	

HUD film is required.

3. Flat Turn Effectiveness - Set up inbound to the left-most strafe panel at the specified airspeed; dive angle as desired. When the in-range cue appears on the pipper, simulate a 1/2 second burst, and immediately transfer attack to the second target in line using decoupled flat turn inputs. If the above run is successful, repeat the run again except after the first simulated/actual burst, transfer the attack to the third target in line. Repeat the procedure so as to determine the maximum separation between primary and secondary targets for which it would be feasible to use flat turn to make a second attack on the same pass. If results are good at 400 knots, repeat the procedure at 460 and so on at 520 knots to determine the effects of increased speed on secondary attack capability.

Table 9.1-8 FLAT TURN EFFECTIVENESS

RUN NO.	KCAS	DIVE ANGLE	DFCS MODE	FCC MODE	ACCOMPLISHED FLIGHT NO.
EV080	400	A/R	SASG	STRF	
081	460	│	│	│	
082	520	A/R	SASG	STRF	

The use of coupled or decoupled flat turn shall be per pilot discretion. HUD film is required.

4. Strafe Pitch Pointing - Set up inbound to the target at the conditions specified in the run table. Prior to arriving within gun range of the target, command, and hold in full nose up pitch pointing. The emphasis of this run is on pitch pointing authority, rates, and gains; therefore, the run may be made by flying perpendicular to a straight narrow road and keeping the pipper on that road during closing by using only pitch pointing inputs. Since the aircraft should automatically trim to the initial dive angle, pitch stick inputs should be minimized to purify the pitch pointing evaluation. The pilot should evaluate the ability to null out elevation errors using left hand controller pitch pointing inputs. Post flight analysis will be used to determine more subtle benefits that may be realized in the strafing task (i.e., increased range from target at first firing opportunity due to nose up pointing, or increased altitude clearance from terrain due to nose down pointing, etc.).

Table 9.1-9 STRAFE PITCH POINTING

RUN NO.	KCAS	DIVE ANGLE	AGL ALT	DFCS MODE	FCC MODE
EV090	400	10°	N/A	DASG	STRF
091	460				
092	520				
093	400	15°			
094	460				
095	520				
096	400	20°			
097	460				
098	520				
099	400	0°	100'		
100	460				
101	520				
102	400		300'		
103	460				
104	520				
105	400		500'		
106	460				
107	520	0°	500'	DASG	STRF

HUD film and C-band tracking is required.

9.1.2.3 Air-to-Air Gunnery. Runs defined in this section are in a slightly different format than those of previous sections. Rather than specify each run on a point by point basis, the target flight conditions and geometry, mode selection, decoupled option use, and tracking tasks will be selected by the JTF at the time of the test. As in all of the evaluation testing, runs/maneuvers other than those currently suggested in this section will be permitted. It shall be the JTF's responsibility to verify that the modified runs are in keeping with the technical and safety objectives contained herein. If necessary, rules of engagement will be briefed before flight to allow for a less structured air-to-air gunnery evaluation.

9.1.2.3.1 Target Maneuvers. The set up conditions for all air-to-air gunnery runs shall be the same. The target shall initially be straight and level at the specified airspeed. The AFTI/F-16 (fighter) will initially be 1NM in trail co-speed. The fighter shall then accelerate to a 50 knot closure rate, and at 3500' range, call for the target to begin his maneuver.

1. Constant g Turn - On command, the target shall begin a constant speed/g/altitude turn at the conditions specified in the run table. Hold the turn for 45 seconds, call "terminate" and roll out.

2. Wind Up Turn - On command, the target shall begin a constant speed/altitude WUT at a build-up rate of 1g per 5 seconds. The build-up should be continued until the target reaches its manpower sustained turn capability at the specified speed. The target shall continue the turn at this load factor for another 5 seconds, call "terminate", and roll out.

3. Turn Reversals - On command the target shall begin a constant speed/g/altitude turn at the condition specified in the run table. The turn shall be held until the fighter calls "reverse". The target shall briskly perform an unloaded roll and reverse the direction of turn. The fighter shall command three reversals for a total of four acquisition/tracking opportunities.

4. Cine Track - On command, the target shall perform a 2g wings level pullup. At 20° pitch attitude, begin a roll to the left while maintaining 2g's so as to arrive at 90° of bank with 45° of pitch attitude. (Power should be increased so as to maintain speed and load factor throughout the high pitch attitude portions of the maneuver.) Continue the left roll to 135° angle of bank. When the pitch attitude reaches 45° nose low, begin a roll to the right so as to arrive at 90° of bank when the nose is 60° below the horizon. (The load factor should be increased to as much as 4g's to keep speed constant.) Continue the right roll so as to reach a wings level state when the nose is on the horizon. From this point, smoothly transition to a low roll rate, 2.5 to 3.0g barrel roll. The maneuver is complete when the wings are level.

9.1.2.3.2 Fighter Tracking Task. Tracking tasks are keyed to evaluation of either gross acquisition or fine tracking capabilities. Any of the three target maneuvers described above are suitable for use with tracking techniques defined below. In all cases, the pilot should attempt to place the pipper within that area of the targets fuselage between the left and right wing roots.

While tracking, range to target should be between 1500 and 3000 feet.

a. Fine Tracking - Immediately upon the onset of the targets maneuver, the pilot should fly so as to maintain the pipper on the target within the specified range.

b. Gross Acquisition - After calling for the target to maneuver, the pilot shall delay his turn towards the target. This delay should be of duration sufficient to force the fighter to perform a high pitch rate maneuver to acquire the target. The pilot should then transition to fine tracking. After establishing a steady state track on the target, the pilot may, if desired, intentionally let the pipper fall off the target and repeat the gross acquisition task.

9.1.2.3.3 DFCS Options. As in the previous paragraphs, the mode selections detailed below are strictly options to be chosen from. They are somewhat listed in the desired order of testing, however, it is not necessary to adhere to this guideline.

a. SAAG Conventional - Use only pitch and roll stick inputs in the tracking task. (No rudder pedal inputs).

b. SAAG Flat Turn - Use flat turn inputs as much as possible tc null out azimuth pipper errors, and pitch stick inputs for elevation errors.

c. DAAG Conventional - Same as SAAG Conventional.

d. DAAG Pointing - Use pointing inputs to null out both elevation and azimuth pipper errors once established in the fine tracking phase.

e. SNRM Conventional - Same as SAAG Conventional.

APPENDIX E

4. DATA ACQUISITION SYSTEM

The Instrumentation Systems provided on the AFTI/F-16 airplane includes the Airborne Data Acquisition System, Airborne Video System and the Camera Systems. The equipment associated with these systems is located in three (3) areas of the airplane as shown in Figure 4-1.

4.1 THE AIRBORNE DATA ACQUISITION SYSTEM

The Airborne Data Acquisition System consists of the following subsystems:

1. Pulse Code Modulation (PCM)
2. Frequency Modulation (FM)
3. Time Correlation (IRIG "B")
4. Telemetry (T/M)
5. Avionics Multiplexed Data Bus (AMUX)
6. Recording.

These subsystems are shown as part of the system block diagram in Figure 4-2.

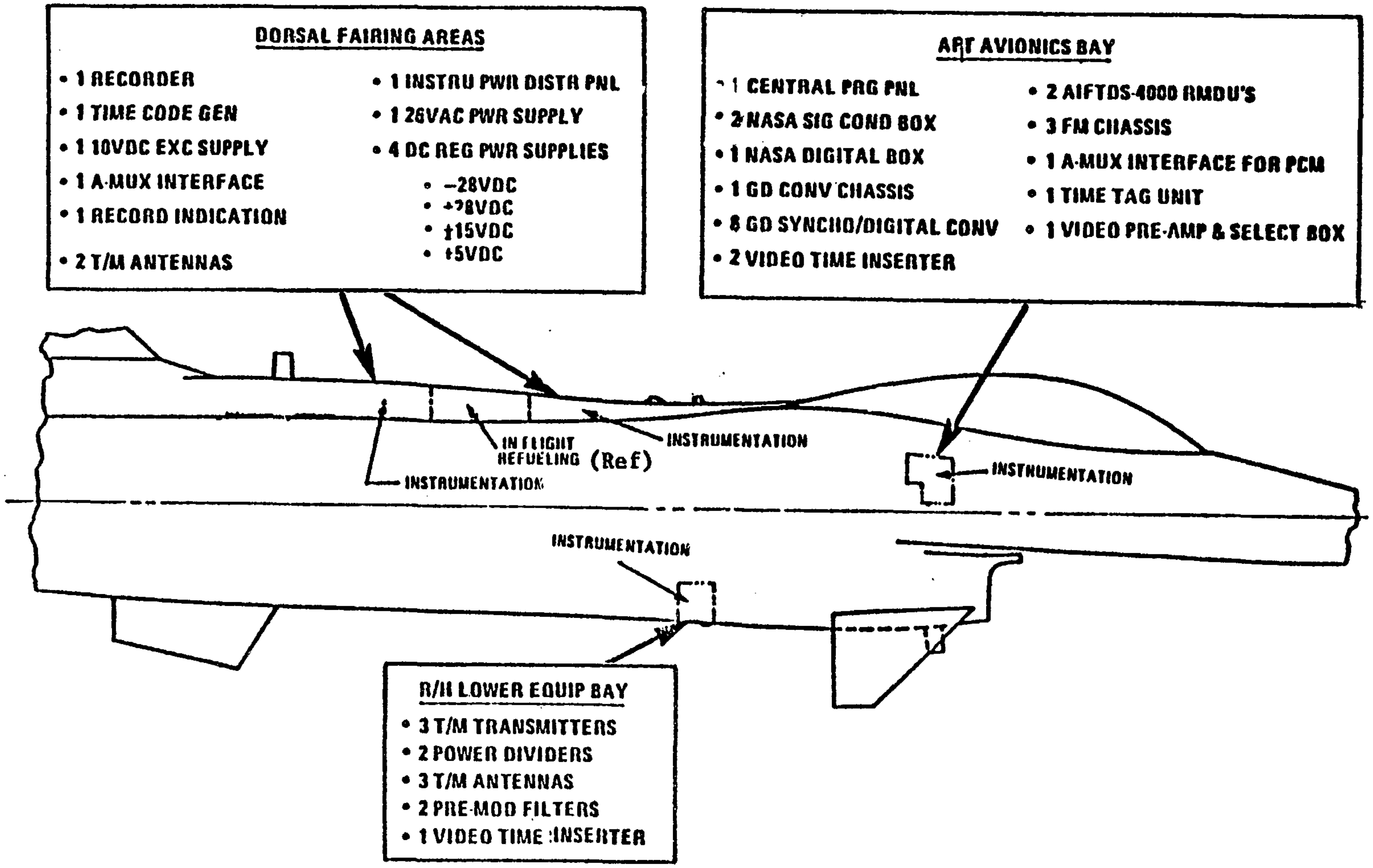
4.1.1 Basic Airplane Instrumentation Provisions

The AFTI/F-16 airplane has instrumentation provisions that include basic harnesses, a data programming panel, and a power distribution panel. The basic harness is routed from remote airplane areas to the main programming panel for the airborne data system. Power is distributed from a central circuit breaker panel which is serviced by the airplane buses. The basic harnesses are utilized for transferring power, control functions, and measurement signals throughout the aircraft. Most of the provisions to accommodate the instrumentation equipment are located in the dorsal fairing area, aft avionics bay, and the lower right equipment bay.

The airplane is equipped with an instrumented test nose boom that replaces the airplanes normal boom and provides angle of attack and angle of sideslip measurement capability.

4.1.2 Pulse Code Modulation

The PCM subsystem uses two (2) Teledyne Controls Model AIFTDS-4000 Remote Multiplex/Demultiplex Units (RMDU). One will have the capacity to format and record up to 200 hardwired measurements and the other will handle up to 255 inserted AMUX data words. Output PCM serial signals will be transmitted to a ground station in addition to being recorded on the test vehicle.



1. AFT AVIONICS BAY WILL CONTAIN THE CENTRAL PROGRAMING PANEL AND SIGNAL CONDITION EQUIPMENT
2. DORSAL FAIRING AREAS WILL CONTAIN TIMING, RECORDING AND POWER DISTRIBUTION EQUIPMENT
3. R/H LOWER EQUIPMENT BAY WILL CONTAIN TELEMETRY EQUIPMENT

Figure 4-1 AFTI/F-16 Instrumentation Equipment

A13697

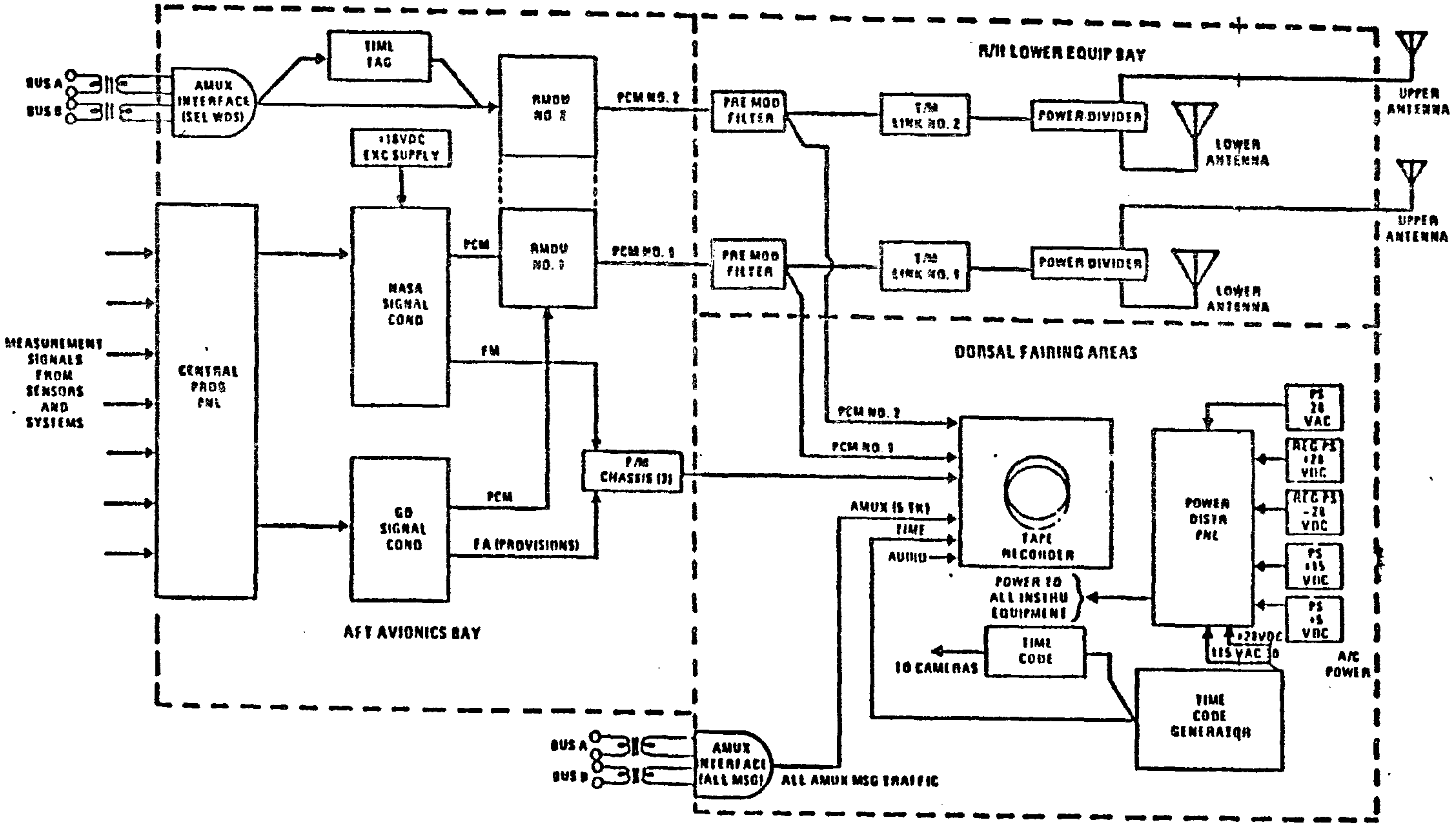


Figure 4-2 AFTI/F-16 Instrumentation System Block Diagram

4.1.3 Frequency Modulation

The FM subsystem is used to record dynamic data where frequency response requirements exceed those of the PCM system. The FM subsystem uses constant-bandwidth and wide-bandwidth IRIG (Inter-Range Instrumentation Group) voltage controlled oscillators and can record up to nine measurements.

4.1.4 Recorder

The on-board data acquisition system uses a 14-track, IRIG, standard, 1-inch magnetic-tape recorder (Ampex AR-700). At a recording speed of 15 inches per second the recorder has a frequency response of 250KHz. The recorder uses a 12 1/2-inch reel of magnetic tape to provide 1 1/2 hours of recording time. The recorder stores data from PCM, FM, AMUX data buses, time code generator, and voice communications.

4.1.5 Time Correlation

The time correlation subsystem consists of an IRIG "B" Time Code Generator and a cockpit time display unit. The time code generator is synchronized with the ground station range master clock prior to each flight of the test vehicle. The on-airplane time code generator provides a common time base for the on-board recorders as a modulated 1 KHz timing signal for tape, parallel output signals for cockpit time display, and a 100-pulse-per-second (PPS) timing for camera film recording. The ground station master clock data is recorded with the telemetry transmissions for time correlation with on-board recorders and cameras.

4.1.6 Telemetry

Three (3) telemetry links are installed on the AFTI/F-16 flight test airplane. Two (2) links are used for PCM and one (1) link is used to transmit selected video signals from the airplane. The first link transmits the PCM data from hardwired aircraft measurements and the second link transmits the PCM associated with selected Avionics Data Bus (AMUX) words along with time tag information. These PCM telemetry links transmit data to a ground facility for use as (1) the Primary Data Source and for (2) "Safety of Flight" Real Time Monitoring. The third link will be used for wideband video signal transmission. This link will transmit the selected video from the CTVS camera or one of the two (2) multipurpose displays in the cockpit. IRIG "B" Time is inserted into the selected video signal prior to being transmitted to the ground station.

4.1.7 Avionics Data Bus (AMUX) Interface Systems

The AFTI/F-16 Airplane Avionics Data Bus (Ref. MIL-STD-1553) will be monitored with two types of interface assemblies as described in the following paragraphs.

4.1.7.1 Total Bus Recording. An interface unit (IH0817) developed and used throughout the F-16 full scale development (FSD) program is used to record the total AMUX data traffic. This unit accepts the single MIL-STD-1553 data stream, inserts bits for word and bus identification, provides synchronization words, and reformats the data into five streams for recording at reduced bandwidth.

4.1.7.2 Interface for Recording Selected Data Words. The AMUX data bus interface subsystem (IH0917) consists of two (2) components. One component will provide programming capability to select messages of interest from the data bus, store these messages in a memory, and sequence the words of the message into preselected positions in the PCM data stream. The second component will provide the necessary "hand shake" operation to insert words, output from the first unit into the PCM data stream. This subsystem provides the capability to merge selected data words from the AMUX data bus with the AIFTDS-4000 PCM System.

4.1.7.3 Time Tag Generator. Data on the MIL-STD-1553 bus monitored in the test program must be time correlated within one millisecond of the actual occurrence. A varying time lag will exist between an event and its insertion in the PCM data stream because of data sampling rates, processing and subcommutation of the selected words from the MIL-STD-1553 data bus. A special time tag generator (IH0916), with the IH0917, provides the means to account for time from an event until the data words, are inserted into the PCM system data stream.

4.1.7.4 Data Pump. Internal parameters in the DFCS and SMS which are not normally outputted on the AMUX can be accessed for instrumentation purposes by placing them in special AMUX blocks called data pump. Any software accessible parameter can be programmed to be included in the data pump block. These parameters are then accessible by the instrumentation system through the AMUX interface system described in paragraph 4.1.7.2. The DFCS and SMS each receive data addresses from the FCC of all parameters to be included on data pump. Data pump can be reprogrammed by altering the FCC OFP. Since the FCC has core memory, OFP changes can be made by loading in a new tape. Data pump address changes can be made as as addendum to the regular FCC OFP tape.

4.2 AIRBORNE VIDEO SYSTEM

The major components of the Airborne Video System (AVS) are the Airborne Video Tape Recorder (AVTR), a modified Cockpit Television Sensor (CTVS) and a Video Select Panel (VSP). The baseline avionics system is designed to route all video signals (i.e., radar, sensor, weapon, or threat), except for the CTVS video, into the Programmable Display Generator (PDG) for display on the Multi-Purpose Displays (MPD). The video signal outputs from each PDG and the CTVS are routed to a Video Signal Select unit as shown in Figure 4-3. The Video Signal Select unit is remotely controlled by the Video Select Panel. One three-position switch on

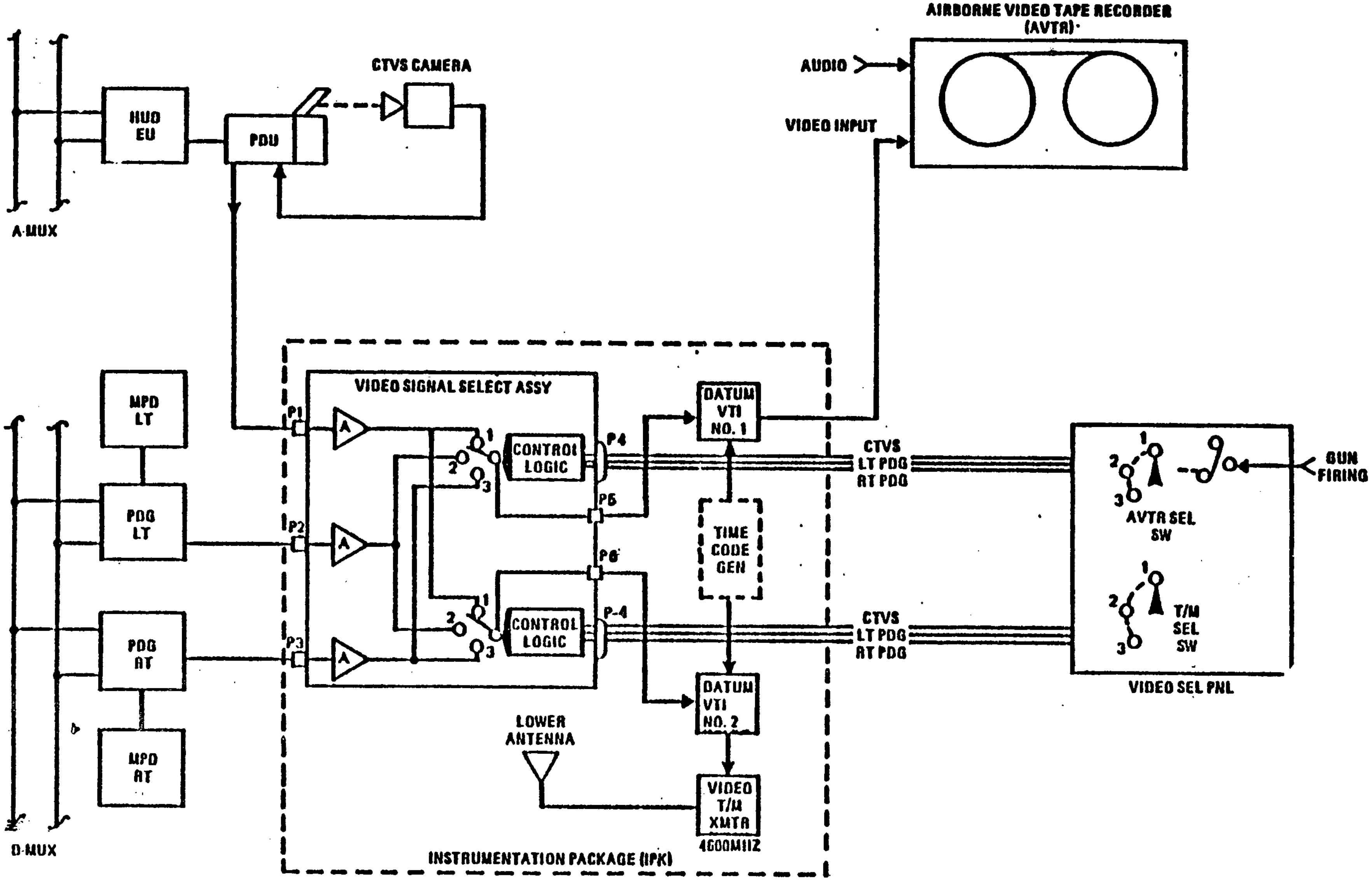


Figure 4-3
Airborne Video System

the Video Select Panel is used to select video (i.e., CTVS, Lt MPD, or RtMPD) for the Airborne Video Tape Recorder (AVTR), and an identical switch on the Video Select Panel is used to select the video signal to be transmitted on the telemetry downlink.

The Video Select Panel has a feature to override the selected AVTR video signal with CTVS video. If the HUD override switch is enabled, the CTVS video is automatically switched into the AVTR record line when the gun is firing (real or simulated). An event marker is generated by the CTVS during gunfire and weapons release.

Prior to recording on the AVTR and transmission on the telemetry downlink, the selected video signals are merged with IRIG "B" time using onboard Video Time Insertion (VTI) units. The inserted time on the video is visible on the upper part of the video picture, reference Figure 4-4. The inserted time has a resolution to one millisecond.

4.3 INSTRUMENTED NOSE BOOM

The instrumented test nose boom on the AFTI/F-16 test vehicle is interchangeable with the aircraft's production boom. It consists of a specially compensated pitot probe and vane assemblies for sensing angle of attack and sideslip. Normal production position error corrections are removed when the test boom is installed.

4.4 CAMERA SYSTEMS

4.4.1 Weapons Separation Cameras

The weapons separation camera system is composed of eight (8) remote cameras, crew station controls, and indicator lights. The cameras record time from the time code generator on the edge of the film. The camera control panel inserts a delay to the weapons release command, at the pilot's discretion, to assure cameras are up to speed at the time weapons are released. There are eight active camera stations for any weapons separation flight. Two cameras are located in each wing tip dummy AIM-9 missile. One camera is located in each strake area looking aft. One camera is located in each chaff/flare dispenser area looking forward. The cameras use 16mm film and operate at 200 frames per second during weapons separation. (See section 8.7.4).

4.4.2 Cockpit Camera

A bracket is provided in the canopy, behind the pilot's seat, for the installation of a forward or aft viewing cockpit camera. The control switch for the cockpit camera is located on the left side of the pilot's panel. The forward viewing, over the pilot's left shoulder, camera is designed for forward external viewing and does not provide coverage of controls or instruments. The aft viewing camera configuration is designed to photograph the spin chute during deployment.

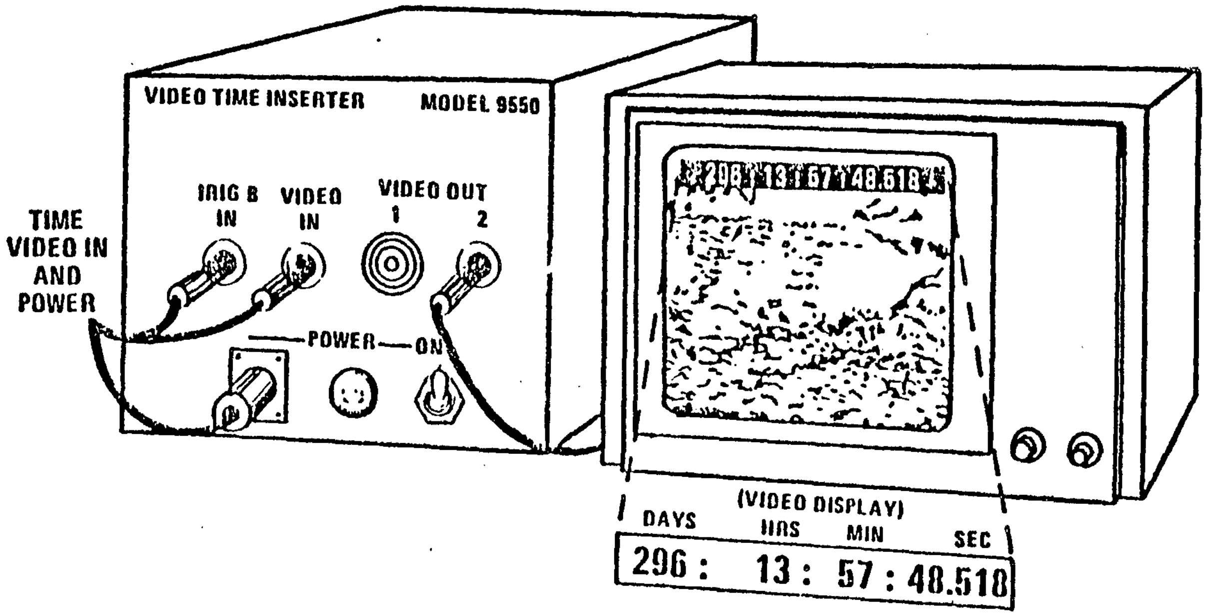


Figure 4-4 Video Time Insertion Capability

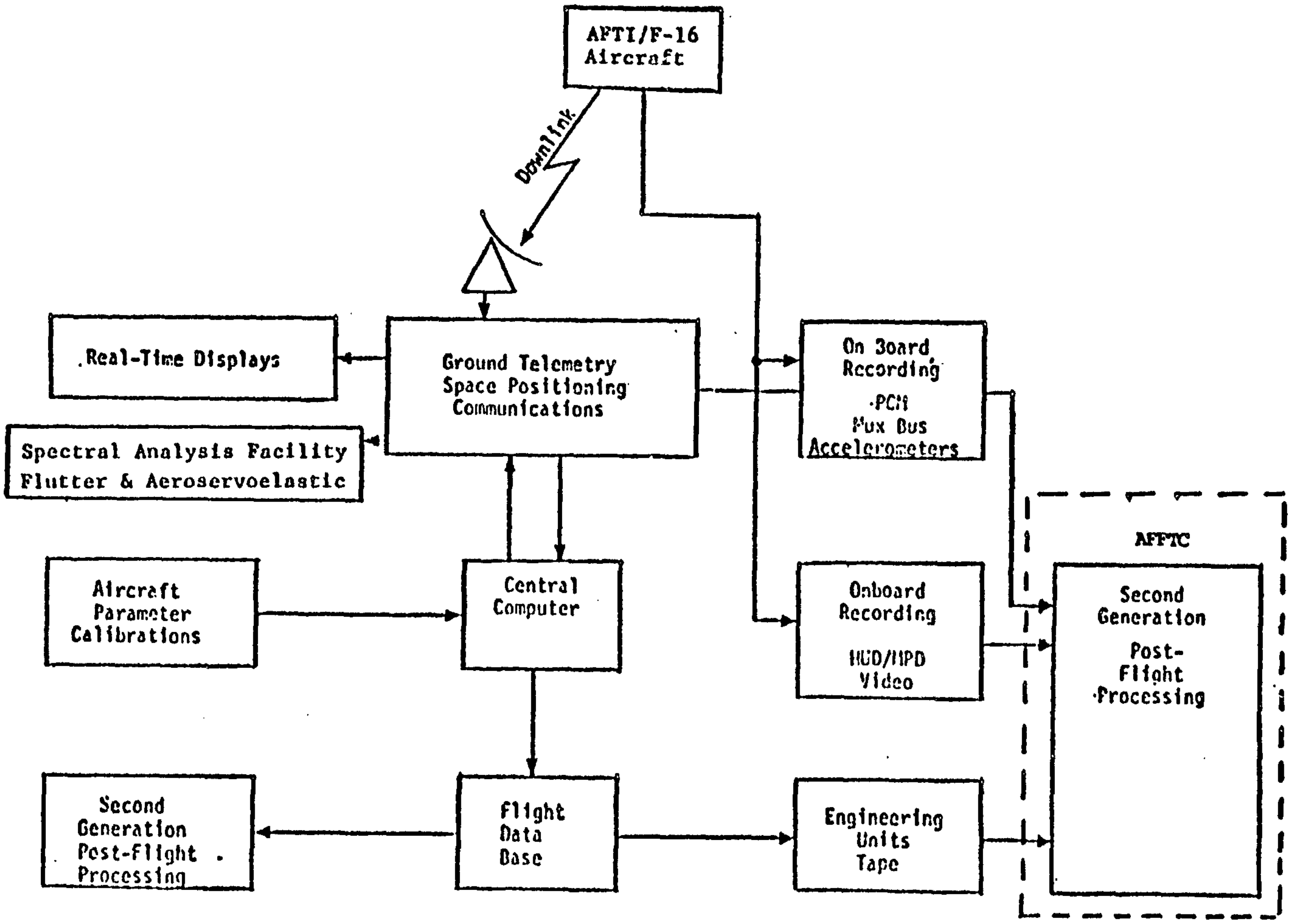


Figure 4-5 AFTI/F-16 Data Flow

4.4.3 HUD Camera

A 16mm film camera is available to replace the Cockpit Television Sensor for selected missions. Installation of this camera calls for relocation of the HUD control panel to a side console. The camera has manual and automatic operating provisions. The automatic mode initiates camera run with the first detent of the gun trigger depressed. Camera aperature adjustment is accomplished manually.

4.5 GROUND TEST EQUIPMENT

Ground support equipment consists of a test cart and calibaration fixtures. These test cart and surface position calibration fixtures are used for pre- and post-flight checkout, calibration, and maintenance of the instrumentation data acquisition system at the airplane.

General Dynamics has modified an existing ground test cart from the F-16 FSD equipment pool to incorporate checkout equipment peculiar to the Teledyne Controls, PCM (AIFTDS-4000) system and the IH0817 and IH0917 AMUX interface assemblies.

4.6 REAL TIME DATA FACILITIES

The NASA DFRC data acquisition and processing system will be used for display of real time processed data and generation of an engineering units tape for AFFTC use in post-flight data processing. A simplified schematic of the ground-based equipment and data flow is shown in Figure 4-5. Data from the aircraft that will be received, processed, and recorded at the ground station includes: continuous UHF transmission of pilot voice, selected video of either MPD or CTVS, PCM stream of hardwired instrumentation parameters, and PCM stream of selected AMUX parameters. During flight, this data will be presented in the form of audio of pilot's voice, engineering unit strip charts, CRT digital displays, and TV monitor display of downlinked video. Additionally, a chornological hard-copy printout of avionic and DFCS discrete events will be provided on an immediate post-flight basis. Early in the program NASA's Spectral Analysis Facility (SAF) will be used to monitor flutter and aeroservoelastic envelope expansion flights and for post-flight data processing.

APPENDIX F

PROPOSED MISSION EFFECTIVENESS SIMULATION DATA PLAN

The following sections describe the type of data to be collected and the analysis of that data.

6.1 DATA COLLECTION

6.1.1 Magnetic Tape

The primary data collection method for this test will be magnetic tape. Only raw data will be collected, with all analysis accomplished off line. The recording format and variable are identical for all runs.

Prior to each data run a set of Header Data will be recorded. This header data will identify the run, and can be used for rapid data sorting by run type or other header variable. Data to be recorded on the header is as follows:

HEADER DATA

1. RUN IDENTIFIER - from the test plan
2. PILOT IDENTIFIER
3. SYSTEM IDENTIFIER
4. TARGET INITIAL LOCATION
5. THREAT LOCATIONS

When the simulator is placed in "GO" the following parameter will be recorded each frame:

FRAME BY FRAME PARAMETERS

1. TIME SINCE "GO"	SECONDS
2. PIPPER AZIMUTH (BODY CO-ORDINATES)	MILLIRADIANS
3. PIPPER ELEVATION (BODY CO-ORDINATES)	MILLIRADIANS
4. FPM AZIMUTH (BODY CO-ORDINATES)	MILLIRADIANS
5. FPM ELEVATION (BODY CO-ORDINATES)	MILLIRADIANS
7. TD BOX AZIMUTH (BODY CO-ORDINATES)	MILLIRADIANS
8. TD BOX ELEVATION (BODY CO-ORDINATES)	MILLIRADIANS
9. TARGET AZIMUTH (BODY CO-ORDINATES)	MILLIRADIANS
10. TARGET ELEVATION (BODY CO-ORDINATES)	MILLIRADIANS
11. DIVE TOSS STEERING DEVIATION	MILLIRADIANS
12. TRIGGER DISCRETE	
13. WEAPON RELEASE DISCRETE	
14. DESIGNATE DISCRETE	
15. LATITUDE	DEGREES
16. LONGITUDE	DEGREES
17. ALTITUDE	FEET
18. YAW	DEGREES
19. PITCH	DEGREES
20. ROLL	DEGREES
21. ALPHA	DEGREES
22. BETA	DEGREES

FRAME BY FRAME PARAMETERS

23. Nz	"g"
24. Ny	"g"
25. AIRSPEED	KTAS
26. PEDAL FORCE	LBS
27. THROTTLE ROTATION	DEGREES
28. STICK PITCH FORCE	LBS
29. STICK ROLL FORCE	LBS
30. THROTTLE POSITION	%
31. TARGET RANGE	FEET
32. BOMB BUTTON DISCRETE	
33. FCS MODE	
34. SMS MODE	
35. DELIVERY MODE (FCC)	
36. MISSION PHASE	

37. TARGET LATITUDE (A-A)
38. TARGET LONGITUDE (A-A)
39. TARGET ALTITUDE (A-A)
40. TARGET YAW (A-A)
41. TARGET PITCH (A-A)
42. TARGET ROLL (A-A)

6.1.2 "Merit" Output

Hard copy print out of the "merit" routine will be available after each data run. This output will serve as a de-briefing tool only. All later analysis will be conducted using raw data from the magnetic tape. The following data will be available from "Merit."

1. TOTAL PIPPER ERROR at release (time average for gunnery)

2. AZIMUTH PIPPER ERROR at release (time average for gunnery)

3. ELEVATION PIPPER ERROR at release (time average for gunnery)

4. BURST LENGTH

5. RANGE at open fire

6. TIME from "GO" to X milliradians total pipper error

7. TIME from "GO" to Y milliradians total pipper error

8. TIME the pipper is within 3, 5, 10, 25, 50, and 100 milliradians (accumulation can begin at "GO", at a constant interval after "GO", or trigger off of 7 above).

6.1.3 Strip Charts

Strip charts will be used to monitor the quality of the testing and to trouble shoot any problem with the simulation. They can be made available in de-briefing if required. Since all data on the strip charts will also be recorded on magnetic tape we do not plan to maintain a file of strip chart data. A list of strip chart variable to be recorded is as follows:

[TBD]

6.1.4 Pilot Questionnaires

Pilots will be asked to fill out questionnaires on selected test runs to collect qualitative data and pilot inputs concerning the cockpit controllers, cockpit displays, general impression of work load, and flying qualities.

The following questionnaires will be used:

[TBD]

6.2 DATA ANALYSIS

The magnetic tapes of raw data will be used as inputs to various figure-of-merit programs and the results will be statistically analyzed with respect to sample variance, mean, and statistical significance. Results will generally be presented in the form of histograms. Conclusions from pilot questionnaires will be summarized, and where application questions will also be summarized in the form of histograms. The following sections describe the figures-of-merit planned for each experiment.

6.2.1 Air-to-Surface Bombing

6.2.1.1 CCIP Alignment Time Experiment. The following figures-of-merit will be utilized for this experiment:

1. Time required to stabilize the Bomb Fall Line (BFL) within ± 2 milliradians of the desired target.

2. Time from BFL stabilization until weapon release.

3. Pipper errors at weapon release in milliradians (total, azimuth, and elevation).

6.2.1.2 CCIP Accuracy Experiment. The following figures-of-merit will be utilized for this experiment:

1. Time from bank angle less than 10° until weapon release.

2. Pipper errors at weapon release in milliradians (total, azimuth, and elevation).

3. Flight conditions at release (dive angle, bank angle, airspeed, altitude, and slant range).

4. Bomb miss distance (selected runs)

5. Probability of Survival (selected runs)

6.2.1.3 Dive Toss Accuracy Experiment. The following figures-of-merit will be utilized for this experiment:

1. Time from bank angle less than 10° until "designate".

2. Target Designator (TD) box errors at "designate" (total, azimuth, and elevation).

3. Azimuth steering errors at weapon release in milliradians.

4. Flight conditions at "designate" (dive angle, bank angle, airspeed, altitude, and slant range).

5. Flight conditions at weapon release (dive angle, and slant range).

6. Bomb miss distance (selected runs)

7. Probability of Survival (selected runs)

6.2.1.4 Turning CCIP Weapon Release. The following figures-of-merit will be utilized for this experiment:

1. Pipper errors at release (total, azimuth, and elevation).

2. Flight conditions at release (dive angle, bank angle, airspeed, altitude, and slant range).

3. Bomb miss distance (selected runs).

4. Probability of Survival (selected runs).

6.2.1.5 Operational System Evaluation. The following figures-of-merit will be utilized for this experiment:

1. Pipper errors at release (total, azimuth, and elevation).

2. Flight conditions at release (dive angle, bank angle, airspeed, altitude, slant range).

6.2.2 Air-to-Surface Strafe

6.2.2.1 CCIP Strafe Accuracy Experiment. The following figures-of-merit will be utilized for this experiment:

1. Time from bank angle less than 10° to "open fire".
2. Flight conditions at "open fire" (dive angle, bank angle, airspeed, altitude, and slant range).
3. Average or RMS pipper error during firing (azimuth, elevation, and total).
4. Average or RMS bullet miss distance (azimuth, elevation, and total-selected runs).
5. Expected number of hits (selected runs).
6. Expected hits divided by total rounds fired (selected runs).
7. Probability of Survival (selected runs).
8. Percent of run time when tracking errors were less than 3, 5, and 10 milliradians.

6.2.2.1 <u>CCIP Strafe Down Pointing Experiment.</u> The following figures-of-merit will be utilized for this equipment:

1. Average or RMS pipper errors (azimuth, elevation, and total)
2. "Open Fire" flight conditions (slant range, altitude, airspeed, dive angle)
3. "Cease Fire" flight conditions
4. Minimum altitude during pass
5. Burst length

6.2.3 Air-to-Air Gunnery

The following figures-of-merit will be utilized for the air-to-air experiments:

1. Average or RMS pipper error during each firing burst (total, azimuth, and elevation)
2. Burst length
3. "Open Fire" range
4. Percent of run time when tracking errors were less than 3, 5, and 10 milliradians.
5. Expected number of hits per run (selected runs)

6. Probability of kill per run
(selected runs)

7. Average or RMS bullet miss distance
(selected runs)

8. Expected number of hits per run divided by total rounds
fired that run
(selected runs)

AFFTC EXPERIENCE WITH PIPPER ERROR AS A HANDLING QUALITIES INDEX

B. Lyle Schofield and Tom Twisdale
Flight Test Technology Branch, AFFTC

During the workshop we heard several references to the idea of using gunsight pipper error - also called aiming error - as a measure of airplane handling qualities. The AFFTC now has nearly eleven years of extensive experience with handling qualities testing using pilot-in-the-loop target tracking maneuvers. All of our experience indicates that aiming error alone is not a reliable indicator of handling qualities. In fact it can be very misleading.

Figure 1 is a collection of four aiming error (pipper motion) histories with the associated Cooper-Harper rating given by the pilot. Note that the best tracking was done with the airplane that was assigned the worst pilot rating. Also note that the tracking performance with the best pilot rating does not differ much from the tracking performance with the worst pilot rating. Finally, note the dramatic difference in tracking performance for the two cases with the same pilot rating. These two cases were flown by the same pilot, in the same airplane, on the same flight, at the same test conditions, one right after the other! These four time histories illustrate that you cannot assess handling qualities or workload by measuring aiming error alone, and expect consistant results.

Figure 2 is a plot of pilot "workload" versus RMS (root mean square) aiming error with associated pilot ratings. Data are presented for levels 1, 2, and 3 lateral-directional handling qualities. There are no discernible, credible trends in these data. One is not surprised that a good pilot can produce good tracking results with an airplane he rates poorly. He does it by working hard. Neither is one surprised that a good pilot can produce good tracking results with less work when the airplane handles well. There are data points in Figure 2 which reflect these kinds of unsurprising cases. There are also some conflicting cases included in Figure 2. For example, there are data points which show that the pilot tracked as well with a level 2 or 3 airplane as he did with the level 1 airplane, but without expending any extra "work" to do so.

We have put the terms "work" and "workload" in quotation marks because what we have measured is only one indication of physical work or activity expended by the pilot during the task. We do not suggest that is is a complete measure of physical work or even that it is an especially good measure of work or activity. We do wish to point out that neither aiming error nor our measure of "work" correlated well with handling qualities level. Only the pilots' Cooper-Harper rating and qualitative comments consistantly reflected the actual handling qualities level of the airplane.

It is clear to us, based on a large volume of experience, that the mechanisms by which a pilot decides whether he likes an airplane or not are not well understood. Certainly it would be a mistake to rely on the tracking performance data we get from AFFTC handling qualities testing.

The key to evaluating handling qualities still resides with the pilots' ratings and qualitative comments. Pilots have often told us that, in their opinion, the mental part of the tracking task is the most fatiguing, especially when the handling qualities are poor and considerable compensation is required. Right now, the only way to account for this important part of the equation is through the pilots subjective evaluation.

The data presented in Figures 1 and 2 are not unusual or worst-case examples. They are representative of the results of thousands of tracking runs conducted at AFFTC over the past eleven years, using HQDT and SIFT handling qualities test and analysis techniques. (HQDT is an acronym for Handling Qualities During Tracking. SIFT is an acronym for System Identification From Tracking.) HQDT and SIFT were developed at AFFTC and are documented in references 1, 2, and 3.

We would be happy to talk with you about our tracking and handling qualities experience at AFFTC. We can be reached at (805) 277-3779 or Autovon 350-3779.

REFERENCES

1. Schofield, B. Lyle, Twisdale, Thomas R., Kitto, William B., and Ashurt, Tice A., Development of Handling Qualities Testing in the 70's: a New Direction, AGARD Flight Mechanics Panel Meeting on Criteria for Handling Qualities of Military Aircraft, April 1982.

2. Twisdale, Thomas R. and Franklin, David L., Captain, USAF, Air Force Flight Test Center, Edwards Air Force Base, California, 93523, Tracking Test Techniques for Handling Qualities Evaluation, May 1975, AFFTC-TD-75-1.

3. Twisdale, Thomas R. and Ashurst, Tice A., Air Force Flight Test Center, Edwards Air Force Base, California, 93523, System Identification From Tracking (SIFT), a New Technique for Handling Qualities Test and Evaluation (Initial Report), November 1977, AFFTC-TR-77-27.

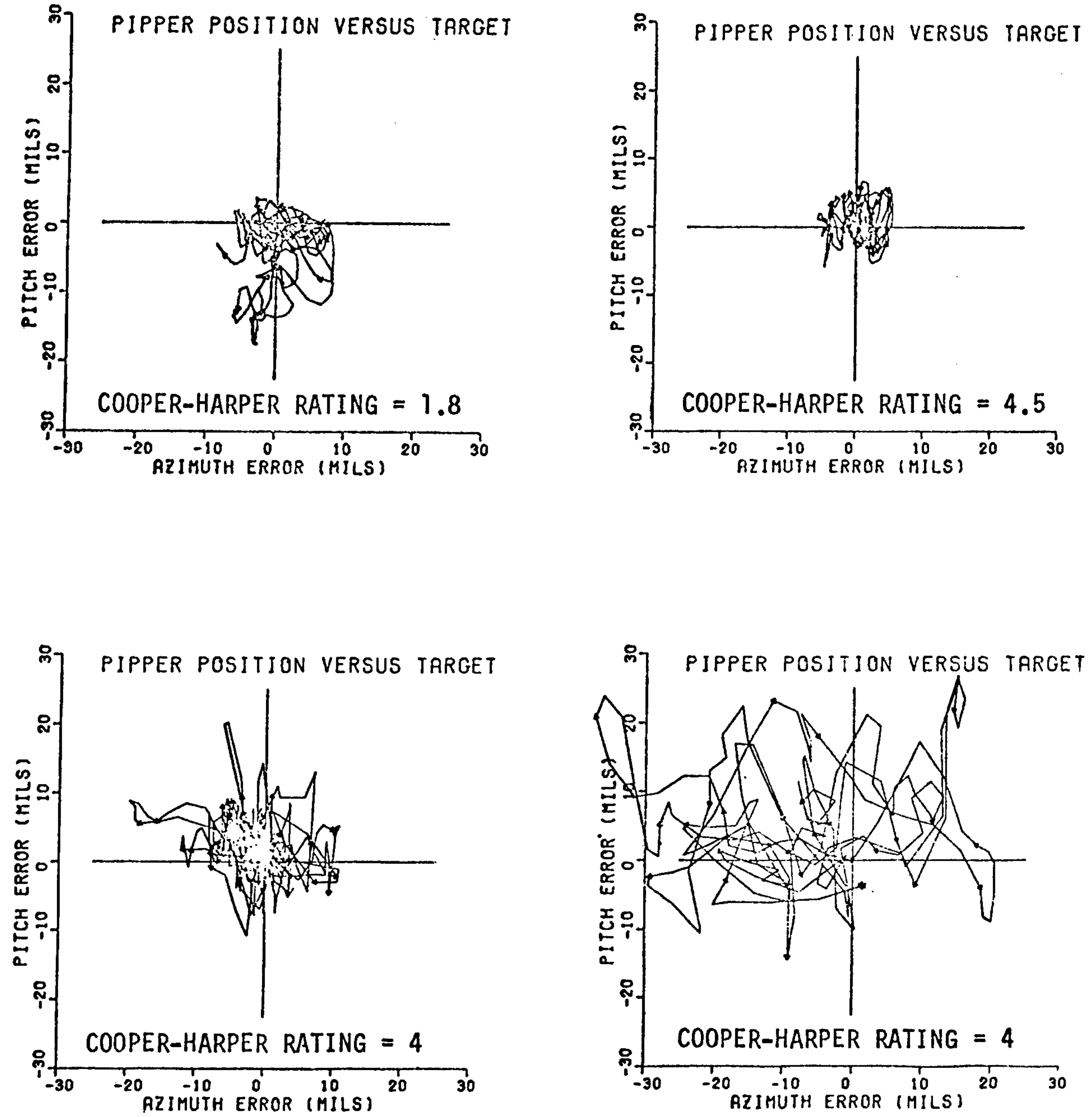


FIGURE 1 EXAMPLES OF TYPICAL SIFT (HQDT) TRACKING RESULTS SHOWING HAZARDS OF RELYING ON TRACKING ERROR AS THE ONLY MEASURE OF HANDLING QUALITIES

$$\text{"work"} = \frac{\text{integration of absolute value of rudder pedal force over duration of tracking run, lb-sec}}{\text{duration of tracking run, sec}}$$

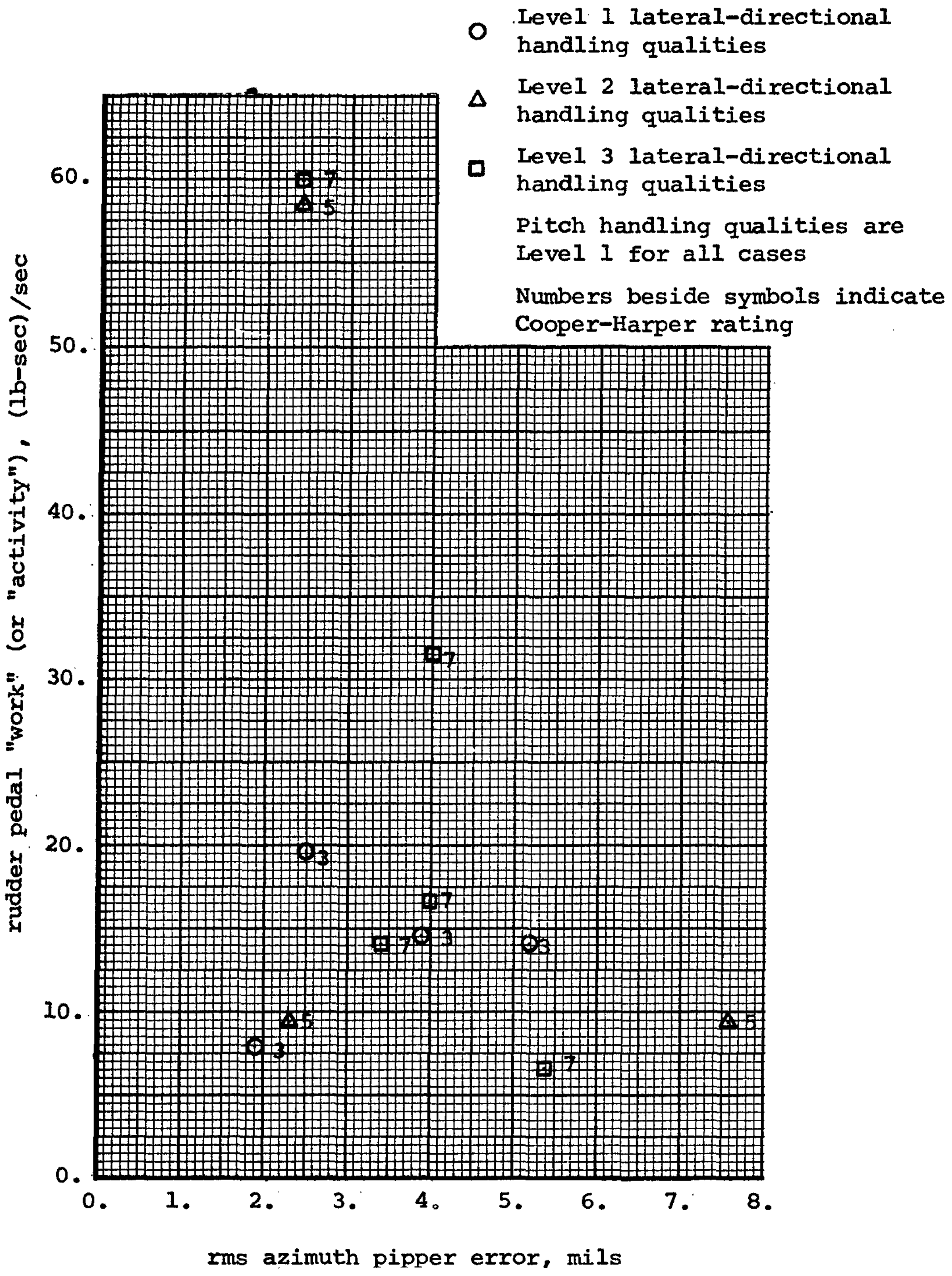


FIGURE 2. Influence of pilot "workload" (or "activity") on tracking error and pilot rating

SUMMARY OF PROCEEDINGS

A workshop was held at Edwards AFB that brought together technical experts in the disciplines of human factors, pilot dynamics, pilot modeling, research and development and test and evaluation to address flight testing to identify pilot workload, pilot dynamics and task performance. The state of the art was laid out by the more than 40 presentations delivered by representatives of Air Force, Navy, Army, NASA, FAA, university, private and foreign organizations (see Appendix A). Informal discussions also were beneficial for exchanging information and ideas. Based upon these inputs and written comments of the participants (Appendix B) the editors have formulated the following broad conclusions and recommendations.

Flight Testing to Identify Pilot Workload.

Workload is a multidimensional concept in which the pilot (physiology, perceptions, technique, training), the vehicle (dynamics, controls, displays, subsystems), the tasks (number, difficulty, relative importance) and the environment (stress, disturbances) all play significant and interrelating roles. Each aspect must be carefully considered in order to effectively assess pilot workload in flight. One or several of these aspects of pilot workload have been objectively measured in flight by Schiflett, Van de Graaff, Roscoe, the Navy Pacific Missile Test Center and the Air Force School of Aerospace Medicine. Important measures that show near-term promise for assessing pilot workload in flight are pilot subjective ratings, rate of pilot control activity, heart rate, and secondary task performance. Several other measures were proposed in the proceedings. The most promising of these for further development appears to be the event-related brain potential described by Donchin and Biferno.

Workload should be specifically addressed throughout the systems acquisition process. Workload technology promises to become as useful in the design, development, test and evaluation of new systems as flying qualities technology is today. Resources should continue to be allocated to measureing pilot workload because increasingly complex mission demands continue to be made of pilots and their aircraft.

Flight Testing to Identify Pilot Dynamics and Task Performance.

These areas are complementary to the evaluation of pilot workload and thus deserve careful attention. Van de Graaff has successfully combined measures of pilot dynamics, pilot workload, and task performance on a helicopter in-flight experiment. Complementary measures for all these areas are listed in an annotated bibliography of "Pilot Performance" measures described by Mixon and Moroney.

Models of pilot dynamics are presently being used to design the dynamic characteristics of flight controls and displays. However, the usefulness of the models has been hampered by the lack of validating flight test data. Methods for identifying pilot dynamics have been successfully used during simulation but in-flight experience is lacking. Often the cues sensed by the pilot are difficult to instrument or measure. However, it is still important to determine the strategy and dynamics used by pilots during critical flying tasks to validate simulator fidelity and aid in vehicle design.

Task performance or mission effectiveness measures are important in the design and evaluation of aircraft. They are quite sensitive to variations in the task and initial conditions and must be applied and interpreted very carefully.

RECOMMENDATIONS.

Two primary technical recommendations for the research and development community resulted from the workshop.

1. Define and introduce into common use a standardized set of objective and subjective workload and task performance measures and procedures. Such a set will encourage those conducting experiments to plan and report findings that will be meaningful throughout the pilot workload, pilot dynamics, and task performance disciplines. Not all the elements of the set would need to be measured in every experiment. Flight test engineers could then adapt the set to particular experimental circumstances.

2. Hold periodic conferences to stimulate the exchange of technical results among the disciplines involved in the measurement of pilot workload, pilot dynamics and mission effectiveness. These conferences will allow the lessons learned in the present workshop to be applied, extended and distributed widely. Perhaps next year's Annual Conference on Manual Control could be expanded to meet this need.

In addition, flight research projects should be undertaken to validate pilot workload measurement techniques and models of pilot dynamics. This is needed to develop confidence in those measurement techniques and models so that they may be effectively used by designers of advanced aircraft.

APPENDIX A

ORGANIZATIONS REPRESENTED

Attending the workshop were 152 people from diverse technical backgrounds and representing 55 organizations. These government, private and foreign organizations are listed here to allow the reader to appreciate the scope and importance of pilot workload and pilot dynamics measurement. Names and addresses of participants will be kept on file by the organizers at the Air Force Flight Test Center, 6520 Test Group/ENAH, Stop 239, Edwards AFB, CA 93523.

GOVERNMENT

U.S. AIR FORCE

HQ Air Force Test and Evaluation Center, Kirtland AFB, NM,

AFWAL/Flight Dynamics Laboratory, Wright-Patterson AFB, OH

Air Force Human Resources Laboratory, Williams AFB, AZ and Brooks AFB, TX

Air Force Academy/DFBL, Colorado Springs, CO

Air Force Aeronautical Systems Division, Wright-Patterson AFB, OH

4950 Test Wing/Wright-Patterson AFB, OH

Air Force School of Aerospace Medicine, Brooks AFB, TX and Edwards AFB, CA

Air Force Flight Test Center, Edwards AFB, CA

Air Force Test Pilot School, Edwards AFB, CA

Air Force Office of Scientific Research, Bolling AFB, D.C.

Air Force Aerospace Medical Research Lab, Wright-Patterson AFB, OH

Air Force Systems Command, Andrews AFB, MD

U.S. NAVY

Naval Air Test Center, Patuxent River, MD

Naval Weapons Center, China Lake, CA

Naval Air Systems Command, Washington D.C.

Naval Air Development Center, Warminster, PA

Naval Health Research Center, San Diego, CA

Pacific Missile Test Center, Point Magu, CA

Naval Post Graduate School, Monterey, CA

GOVERNMENT

U.S. ARMY

Army Research & Technology Labs (AVRADCOM), Moffett Field, CA

Army Human Engineering Lab, Aberdeen Proving Grounds, MD

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

NASA Ames Research Center, Moffet Field, CA

NASA Johnson Space Center, Houston, TX

NASA Dryden Flight Research Facility, Edwards AFB, CA

NASA Langley Research Center, Hampton, VA

F.A.A.

Federal Aviation Administration, Northwest Region, Seattle, WA

PRIVATE

CONTRACTORS

Lockheed-Georgia Company, Marietta, GA

Peceptronics Inc., Woodland Hills, CA

Rockwell International, Los Angeles, CA

Bell Labs, Pisataway, NJ

Systems Technology Inc., Mountain View, CA and Hawthorne, CA

Canyon Research Group Inc., West Lake Village, CA

Salisbury Labs, Worchester, MA

Douglas Aircraft, Long Beach, CA

Northrop Corp., Hawthorne, CA

McDonnell Douglas Corp., St Louis, MO

SRI International, Menlo Park, CA

Analytical Mechanics Association, Mountain View, CA

BDM Corp., Dayton, OH

Boeing Computer Service, Seattle, WA

General Dynamics Corp., Ft Worth, TX

Alphatech Inc., Burlington, MA

Systems Control Technology Inc., Palo Alto, CA

Systems Research Labs Inc., Dayton, OH

Flight Engineers International Association, Washington D.C.

Bolt Beranek and Newman Inc., Cambridge, MA

Hughes Helicopters, Culver City, CA

Honeywell Systems Research, Minneapolis, MN

UNIVERSITIES

Arizona State University, Tempe, AZ

University of Illinois, Champaign, IL

University of Southern California, Los Angeles, CA

Purdue University, West Lafayette, IN

Wright State University, Dayton, OH

Worcester Polytechnic Institute, Worcester, MA

FOREIGN

National Aerospace Lab (NLR), Amsterdam, Netherlands

Royal Aeronautical Establishment, Bedford, England

APPENDIX B

RESPONSES OF PARTICIPANTS

On the final day of the workshop, the participants were asked to provide feedback on the workshop by responding to two questions:

1. What were your objectives for this workshop? How well were they accomplished?

2. Where should we go from here?

The written responses have been edited and are published here to allow the reader to appreciate the general consensus on the state of the art of flight testing to identify pilot workload and pilot dynamics.

QUESTION 1:

What were your objectives for this workshop? How well were they accomplished?

1. "My objective was to find out what the state of the art in workload measurement was and what tools were available. To this end my objectives were met."

2. "Would have liked more opportunity for discussion - to define and focus attention on main issues. One objective that was achieved was to at least get aircraft and human factor/psychology types together."

3. "Get overview of recent engineering and psychological approaches. Accomplished to a great degree, due to large sample papers, good attendance."

4. Workshop Objectives of AFTEC Team:

   a. Bring attendees up to the state of the art for their mutual comprehension and use.

   b. Promote further development of methods through stimulation of thought/interaction.

   c. Organize and coordinate workload efforts in design, testing and evaluation on a national basis to optimize the return on all our efforts.

   d. Utilize our voices on an organized basis to promote the proper addressing of human factors issues from the very inception of new systems.

e. Provide scientifically valid/realistic operational/logistical assessments of new systems in terms of their operators. (DOD 5000.3)

5. "Objectives were to try to ascertain state of the art in workload measurement, and significance of workload levels, for application in civil transport certification."

6. "My objectives in attending a conference are usually twofold: (a) to update my store of relevant technical information, and to help others do the same, and (b) make professional and business contacts. The technical interchange was useful, although not quite what I had expected. Specifically, the format was more like that of a standard conference of prepared papers, with little of the kind of give and take that one associates with a workshop. The potential for contacts more than met my expectations."

7. "My objective in attending this workshop was to obtain an overview of the technology used to measure workload and model pilots, to find what the current state of the art has to offer (it really is an art), and to meet those individuals who are at the forefront of this research. The very broad spectrum of presentations which were assembled for the workshop did indeed meet these objectives. I wanted to learn about those tools which have been developed, tested, and validated that I could obtain for immediate use on a Pilots Factors Research contract. Unfortunately, this objective has not been totally fulfilled. While I will obtain and use SWAT from AFAMRL/HE, I had personally hoped that a tested and validated objective workload assessment technique (OWAT?) would be available."

8. "Objective was to understand various facets of the workload problem and to listen to people from diverse backgrounds. Both these objectives were met very well."

9. "My objectives for this workshop were to receive a rapid, efficient update on state of the art in pilot workload. I use workload as a tool in evaluating alternative design approaches in applied settings. My update objective was very much achieved. I was disappointed because I hoped - fantasized - that greater progress would have been revealed. However, I am nonetheless impressed with the amount of attention workload is receiving and am very hopeful for the future."

10. "Attendance: to obtain awareness of current efforts in pilot workload. Comments: excellent conference (a good sound system would have been appreciated)."

11. "My goal was to find out about recent developments in pilot performance measurement and pilot workload measurement. I am very satisfied with the Thursday session which is the only one I could attend."

12. "Everyone (almost) in Workload and Pilot Models came very well organized (albeit casual) to your party. Good instructions and map (should have had motels identified thereon). Good overall plan and durations: Military people with needs and overview of work (day 1), Topic A (day 2), Topic B (day 3) allows specialists to attend on particular days. The name Workshop is a misnomer. A Workshop includes several small groups talking informally and with much more dialogue and interchange. It is closest to a Conference (no drinking) or Symposium (includes wine, women and song, look it up in American Heritage Dictionary). More questions and interchange would be desirable, _if_ a Workshop is intended. _But_, for more coverage in a conference format the idea of presentations and private conferences at long breaks is a viable alternative."

13. "My objectives were twofold: to determine the orientation of the approaches being pursued in comparison to the approach I plan to implement as part of our research program; to obtain a feel for the Air Force desires concerning workload assessment as part of the design and flight test phases in order to suggest to our company management a posture which will permit us to more effectively meet the requirements of our primary customer."

14. "I came to the workshop to (1) share my ideas and get feedback, (2) get more ideas on pilot parameter I.D., (3) hear about the AFTI/F-16 flight test to see how we might better interact or support it. I think it was great, especially the interdisciplinary aspects of it."

QUESTION 2:

Where should we go from here?

1. "Meet at least quarterly in support of the same objectives. Begin a standardization of terminology which appears to be in pretty good shape already. Work to enhance the quality of all methods: task analysis, computer modelling, subjective methods, performance methods. Promote combined methods. Support in our organizations the development of multisensor capable mission area simulator(s) to replace the current fragmented/piecemeal approach and be available to all developers/testers. Support development of airborne test vehicles for advanced manned systems."

2. "A large fraction of human factors related accidents result during low workload levels. I think the reason is related to low workload, monotony, etc. Hopefully, in the future this will be addressed, because the problem is getting worse. With more automation, climb, cruise, and approach phases find the pilot not actively in the control loop, but a monitor and/or manager. I'm not sure he's very good at this. It would seem that a major human factors problem is minimizing error rate output. Workload is only one facet of this."

3. "Have to get funding - sell management. Next time, allow time for questions. Great job by organizers!"

4. "I would still like to see a true workshop on pilot workload, but I don't think the rate of technical progress is sufficiently great to warrant such a specialized activity every year. Let me suggest the following approach: First, encourage researchers to submit their latest results to an appropriate yearly conference (the Annual Manual might be a good one) which makes a point of always having a session on pilot workload. Second, have a periodic but less frequent (say every 2 or 3 years) review of the field in which there is a collective attempt to define the state of the field and to recommend approaches for filling knowledge gaps. Such a review would divide time between formal presentations and roundtable discussion."

5. "Who is "we"? People interested in workload, but different aspects: physiological, objective and subjective measures, models. It would be worthwhile to summarize their positions/progress/status, etc. A subcommittee approach might be useful. Each subcommittee makes presentations to those of like interest and later the group chairman of each group reports out to the larger body. This info is exchanged on broader level. Coordinate with WORKLOAD TAG (Technical Advisory Group). Keep operational people involved."

6. "About the conference - I think it was a good-to-great one time event. We already have Annual Manual, TAG and SUB-TAG groups and interest groups associated with workload at Human Factors meetings. So there are already several Annual or Semiannual meetings at which workload is the prime or only topic. A unique feature of your meeting was that behavioral and control theory types could get together to compare notes. I think you'll have real problems passing the meeting on to others to manage. The location offers an outstanding opportunity to those poor unfortunates who've never been assigned there. I favor a periodic meeting like the one you organized but with enough time in between for new material to emerge, new needs to develop and the ripples from prior meetings to die down. Once every 2 or 3 years would be about right. Also, with a bit

of thinking you could come up with a Thematic variation on workload and its many facets that could enliven interest in the topic. I think the flight test requirement issue will only be carried by folks like you and NATC where it is a daily concern. I encourage you to keep it a part of your meetings."

7. "Every speaker should provide his/her own operational definition of 'workload'."

8. "There needs to be more pooling of knowledge by people from different DoD and civilian agencies and more combined test efforts. Workshops such as this should help. It seems to me that a lot of people are independently working on similar programs and that the field of human factors would benefit from more consolidation of effort."

9. "Generate needs/requirements statement, from technical viewpoint, required RDT&E WL/pilot dynamics funding efforts. Highlight successful integrated efforts tri-service/FAA/NASA as examples for continued and new RDT&E funding efforts. Integrate test requirements, e.g., software into prototype weapon systems for flexible changes required in OT&E and OPTE for fleet data acquisition and analysis. Recommendations should be made through established TAG representatives which have the "charter and blessing" of higher DoD authority for implementation, e.g., specs."

10. "Where do we go from here? (a) continue these workshops on a annual or biannual basis; (b) improve audio/visual conditions; (c) increased emphasis on flight test data and flight test validation of ground based results. Note: This requires increased pressure by DoD agencies interested in pilot workload data; pressure on other DoD agencies that influence who can collect what kind of data in flight test programs (RDT&E and OT&E). Flight Test Data - non-availability: much data, both qualitative and quantitative, is being generated in flight test operations at EAFB and elsewhere. These data are being developed in RDT&E and OT&E types of flight testing. The vast majority of Air Force and Navy agency reports containing these data are limited distribution to U.S. Government agencies only that are involved in test and evaluation. The dissemination of these data are very severely reduced because they are not readily available and require considerable time and effort to obtain. It should be noted that a very large percentage of these limited distribution statements is not based on military security or need to know. Proprietary and other flimsy allegations are given as an excuse for limited distribution of data reports. The cost of development of the flight test data reports starting with the conceptual design of the aircraft system years earlier was paid for by the U.S. taxpayers. This group, U.S. taxpayers, also pay

for the development of follow-on systems whose designers are unable to benefit from the mistakes (as well as good design features) made on earlier systems that have recently undergone flight test."

11. "Include names and addresses of attendees in Proceedings to help people to keep in contact. Get proceedings out quickly."

12. "The modern cockpit will incorporate many new control and display systems. A goal in workload evaluation should be to help in defining what the modern cockpit should look like. We have the technology to do anything. What should we do? Basic pilot modelling has been done using "S" domain models. If one considers modelling the pilot by comparison to a digital computer (three-part) the model would be more adaptable to the variety of processes the human is capable of doing.

13. "In my opinion, eventually we are going to have to develop absolute workload criteria. Essentially, current workload measures are comparitive measures, i.e., which candidate design has the lowest workload. This information is useful to a point, but the real question is not what design is best, instead it is which designs have an 'acceptable' workload. No matter how much better one design is using comparitive measures, there is no guarantee that any of the designs have an 'acceptable" workload. The determination of absolute workload acceptability criteria will be even more time consuming, difficult and frustrating; however, I really think this is where we need to start orienting our efforts because this is where the real benefits of workload assessment lie."

14. "Should have one every two years; 1 in west, 1 in east (Pax R?)."

15. "Conferences like this should probably be held every year. Everyone should be encouraged to measure both workload and performance simultaneously whenever possible."

16. "Suggest consideration of a steering group (Tri-Service plus industry representation)."

17. "Keep it going every one or two years."

18. "In terms of 'where do we go from here', I believe that both the operational severity of the workload problem and the breadth of techniques proposed call for the following activities: (a) formulation of a list of 'action items' to address both short and long term problems in workload assessment/pilot modelling. What needs to be done? When? What is the most pressing issue? Prioritize all issues (by

workshop attendee agreement, by status of current research, by dollars available?); (b) Assign appropriate action items to those individuals or organizations most suited to perform them; (c) standardize techniques, methodologies, handbooks, scenarios, etc., and distribute them! Perhaps a model repository at Air Force Flight Dynamics Lab which could exercise configuration control would be useful for the efficient accomplishment of a structured set of action items in an efficient manner; (d) continue these workshops. Perhaps hold them in conjunction with the Annual Manual to (i) avoid duplication, (ii) increase depth and breadth of papers, (iii) save travel funds. If this is done, the pilot dynamics presentations and the pilot workload presentations could be given over a 4-5 day period; (e) produce a tri-service document which lists people, agencies, projects, schedules, and funds. (Similar to JTCG/AS publication.) This would help to avoid duplication and identify gaps with which to focus future research; (f) develop a state vector for pilot workload. This multi-dimensional metric could be standardized for use throughout DoD/FAA/NASA and might look like:

$$\begin{array}{l}\text{Workload} \\ \text{Metric}\end{array} = \begin{bmatrix} \begin{bmatrix} \text{Physiological Measures} \\ \text{(i.e., heart rate, etc.)} \end{bmatrix} \\ \\ \begin{bmatrix} \text{Performance Measures} \\ \text{(i.e., RMS error, etc.)} \end{bmatrix} \end{bmatrix}$$

If this is developed, then anyone performing research could report their results in a standard format even though not all the cells in the vector might not/could not be measured; (g) use the NTEC document, discussed by Cmdr Moroney of NPGS, as a baseline for future reports and articles. Then new reports should be abstracted with the keywords used in the NTEC report. Perhaps yearly updates to the NTEC report should be planned for and implemented.