(NASA-CR-169917) CDTI: CREW FUNCTION ASSESSMENT Final Report, Jun. 1979 - Dec, [illegible] (Massachusetts Inst. of Tech.) 48 p HC A03/MF A01 CSCL 05H
N83-19471
Unclas
G3/54 02887

FINAL REPORT

CDTI - CREW FUNCTION ASSESSMENT
NASA GRANT #NCC1-23

Harvard/M.I.T. Division of Health Sciences and Technology
to
Flight Operations Research Branch
Flight Control Systems Division
NASA Langley Research Center

December 31, 1982

MAR 1983
RECEIVED
NASA STI FACILITY
ACCESS DEPT.

Period Covered: 6/79 - 12/82

Principal Investigator:
JohnR. Tole, Harvard/M.I.T. Division of HealthSciencesand Technology, MIT

Co-Investigator:
Laurence R. Young, Professor of Aeronautics and Astronautics, MIT

NASA Investigator:
Randall Harris, Sr., Ph.D., Flight Operations Research Branch, NASA Langley Research Center

ORIGINAL PAGE IS
OF POOR QUALITY

ABSTRACT

Sophisticated man-machine interaction often requires the human operator to perform a stereotyped scan of various instruments in order to monitor and/or control a system. For situations in which this type of stereotyped behavior exists, such as certain phases of instrument flight, scan pattern has been shown to be altered by the imposition of simultaneous verbal tasks. This report describes a study designed to examine the relationship between pilot visual scan of instruments and mental workload. It was found that a verbal loading task of varying difficulty causes pilots to stare at the primary instrument as the difficulty increases and to shed looks at instruments of less importance. The verbal loading task also affected the rank ordering of the scanning sequences. By examining the behavior of pilots with widely varying skill levels, it was suggested that these effects occur most strongly at lower skill levels and are less apparent at high skill levels. A graphical interpretation of the hypothetical relationship between skill, workload, and performance is introduced and modelling results are presented to support this interpretation.

In addition a measure of entropy of the scan is introduced and, as a measure of the randomness of the scan, appears to be closely related to the measured verbal task load. In a parallel manner periodicity of the scan, as reflected by its autocorrelation was found to be of particular interest in assessing pilot response to increasing mental workload.

ORIGINAL PAGE IS
OF POOR QUALITY

Acknowledgements

A number of persons have made important contributions to the study described here. A. Ephrath's knowledge of pilot performance and experimental design were of immense value to the project as was his ever present wit.

A. Natapoff suggested the use of the entropy formulation to summarize order in scan patterns. N. Moray introduced the principal investigator to the Wittenborn test which was used as the mental loading task. M. Goode was extremely helpful in the configuration and operation of the oculometer, while A. Spady offered many useful insights and anecdotes on the problems of piloting an aircraft.

We were fortunate to have a large group of enthusiatic and cooperative pilot subjects and are grateful for their contributions to this study. The comments and suggestions of the four NASA test pilots who participated were particularly useful in the ultimate conduct of the experiments.

CONTENTS

Abstract i

Acknowledgements ii

Summary iii

Introduction 1

Importance of Mental Workload 1

Rationale for Studying the Instrument Scan 2

A Conceptual Framework for the Study 3

Experimental Procedure 5

Piloting Task 5

Mental Loading Task 6

Visual Side Task for Workload Calibration 7

Conduct of the Experiments 8

Equipment 8

Independent Estimate of Pilot Skill 9

Results 12

Initial Data Analysis 12

Dwell Time Histograms 13

Fixation Sequences 17

Quantifying Disorder in the Scanpath 18

A Revised Method for Calculating Entropy 22

Entropy Rate 23

Autocorrelation and Power-Spectral Density 25

Performance Measures 29

A Model Relating Workload, Performance, and Skill 31

Concluding Remarks 35

Publications from this Research 37

References 38

Figures

1. Hypothetical Relationship between Performance/Skill/Workload 5

2. Schematic of Precision Straight & Level Flight 6

3. Mental Loading Task Algorithm 7

4. ATC 510 Simulator 10

5. Microcomputer System used for Experimental Control and Data Analysis 11

6. Raw Scanning Data 14

7. Dwell time histograms 15

8/9 % Fixation Sequence usage 19,20

10. Average Entropy Rate vs Task Difficulty 24

11. Autocorrelation Plots for pilot 4 27

12. PSD Plots for pilot 4 28

13. Performance/Skill/Workload Model development data 33

14. Performance/Skill/Workload Model test data 34

Tables

I. Relative Skill Scores of all subjects 12

II. Workload Sidetask Results 13

III. % of Instrument Dwells > 5 sec for Different Loading Levels 17

IV. Dominant Frequency in Scan of Attitude Indicator for all subjects for Different Loading Levels 29

Equations

1. Difficultyof Mental Loading Task as a Function of Interval Between Number Presentation 7

2. Calculation of Workload Index from Visual Side Task 8

3. Estimate of Pilot Skill based on Experience Factors 9

4. Definition of Entropy of a Sequence 21

5. Maximum Number of Possible Fixation Sequences of Length N 22

6. Entropy Normalized for Total Number of Fixations in a Run 22

7. Definition of Entropy Rate 23

8. Practical Expression for Entropy Rate 23

9. Entropy Rate as a Function of Mental Task Difficulty 25

10. Mental Task Difficulty as a Function of Entropy Rate 25

11. Zero Mean Binary Fixation Sequence 25

12. Sample Autocovariance of Fixation Sequence in (11) 26

13. Performance Score For Mental Loading Task 30

14. Performance on Piloting Task 31

15. Model Relating Performance, Skill, and Mental Task Difficulty 31

16. Rearrangement of equation (15) 31

17. Rearrangement of Equation 14 according to equation (16) 32

18. Equation Resulting from Regression on Equation (16) 32

19. Performance based on Regression from Equation (18) 32

20. Normalized Performance 32

ORIGINAL PAGE IS
OF POOR QUALITY

SUMMARY

The experimental method described herein required pilots to maintain a general aviation flight simulator on a straight and level, constant sensitivity, Instrument Landing System (ILS) course with a low level of turbulence. An additional periodic verbal task whose difficulty increased with frequency was used to increment the subject's mental workload. The subject's lookpoint on the instrument panel during each ten minute run was computed via a TV oculometer and stored. Several pilots ranging in skill from novices to test pilots took part in the experiment.

The results indicate an increase in fixation dwell times, especially on the primary instrument, with increased mental loading task. The amount of "staring" observed appears to depend on the level of skill of the pilot; skilled subjects appear to stare less under increased loading than do more 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.

Analysis of the periodicity of the subject's instrument scan was accomplished using autocorrelation. Skilled pilots were found, when stressed, to scan their primary instrument in a periodic fashion. The period was related to the interval between number task presentation. A similar result was not observed in novice pilots. This finding suggests that skilled pilots may handle the additional loading task in a much more systematic fashion that do novice pilots.

Entropy rate (bits/sec) of the sequence of fixations was also used to quantify the scan pattern. It consistently decreased for most subjects over the four loading levels used. 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. This estimate was used to quantify the workload of the subject.

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

The above results suggest the possible utility of instrument scan in the quantification of mental workload and/or pilot skill during constant piloting tasks. Methods were also suggested for studying variations in pilot workload during short epochs, though these have not been attempted as yet.

ORIGINAL PAGE IS
OF POOR QUALITY

INTRODUCTION

This report summarizes research conducted to study the relationship between the instrument scan of an aircraft pilot and the level of difficulty of the several tasks of flying an airplane. The work originally concerned a specific question: the quantitative comparison of the mental workload of conventional cockpit displays vs. novel CRT displays such as the Cockpit Display of Traffic Information (CDTI). However as the study progressed, it became clear that more fundamental work on the nature and quantification of the effects of mental workload on visual scanning behavior was necessary before such a comparison could be made. Thus, the evolution of the research has been away from the specific question first posed and toward developing a basic understanding of visual scanning in pilots and of the interrelationships between the instrument scan and piloting performance, skill, and mental workload.

This work has yielded an experimental paradigm for studying visual scanning behavior, several techniques for quantifying this behavior, and has suggested a number of possible avenues for further research. The techniques developed during the project have been applied to several practical questions in aviation.

Preliminary experiments using the NASA Langley Terminal Configured Vehicle (TCV) simulator with CRT instruments and a Microwave Landing System (MLS) simulation served to help define the requirement of an experimental protocol to study instrument scan and pilot workload while also illustrating the problems in attempting to study complex man-machine interactions.

The final set of experiments described here were conducted using a desktop general aviation simulator. The piloting task involved maintaining this simulator on a straight and level, constant sensitivity, Instrument Landing System (ILS) course with a low level of turbulence. A task employing an algorithm based on relative magnitudes of a sequence of numbers was used to increment the subject's mental workload. The task was presented at periodic intervals which caused the difficulty of the task to increase with increasing frequency of presentation. The level of loading for various conditions was also estimated in an independent series of runs using a side task. The subject's lookpoint on the instrument panel during each ten minute run was computed via an oculometer and stored. A total of thirteen pilots of varying skill participated in two sets of experiments.

Importance of Mental Workload

The desire to measure workload is usually motivated by the need to predict situations in which operator performance will decline. The reasons for this are evident: if the operator has too many tasks to accomplish in too short a time, the performance on all or some of the tasks may be diminished. The same may be

ORIGINAL PAGE IS
OF POOR QUALITY

true if the operator allows his attention to wane because the system he is controlling is highly automated. The latter is termed a condition of underload.

Since a goal of workload measurement is the prediction of performance, it is often suggested that performance is the parameter which should be measured as the loading conditions are varied. Certain performance criteria may be set and when the pilot cannot meet them the level of loading may be judged to be too high. Such a technique assumes that performance varies in a consistent fashion with loading and skill. Thus, for this approach to be generally useful, all pilots should experience about the same performance decrement for the same increase in workload. Experience suggests that this is not the case however. In activities such as piloting (or playing a musical instrument or participating in an athletic event) where the simultaneous conduct of manual dexterity and verbal or mental tasks is especially important, performance of a skilled operator may not show any decrement (or may even improve) until loading is severe, and then a precipitous decline in performance may occur. Since the skill of commercial or test pilots is high, it is difficult to determine subtle differences in workload via performance decrement when they are used as subjects. One goal of this research is a non-invasive measure of workload which does not depend heavily on skill. Some aspects of visual scanning behavior may yield this result.

Rationale for Studying the Instrument Scan

If one hypothesizes that some repetitive piloting task will invoke a regular visual scan (spatial/temporal pattern of eye movements) during instrument flight then it may be possible to observe changes in this scan as external factors such as noise, interruptions or other side tasks, and fatigue interfere with the piloting task. If this hypothesis is correct, then alterations in the scan pattern used by the pilot may be an indicator of either fatigue or increased/decreased mental workload.

The analysis of a subject's visual scan has been examined by various workers in an effort to study behavior. Numerous investigators have studied the patterns of eye movements during the viewing of scenes, pictures, etc. (Noton and Stark, 1971; Senders, 1970; Fisher, et.al., 1981). If a picture is being viewed, it is frequently observed that, after an initial period of general inspection of the scene, the scan tends to return frequently to the points of highest interest to the subject. Ambiguous figures such as the Necker cube (Ellis and Stark, 1978) have been used to determine whether the visual scan provides a clue on the nature of the perceived image. A common feature of these various experiments seems to be the allowance of free eye movements in viewing the target(s). Thus the scan pattern which develops is driven largely by the subject and not by the scene.

ORIGINAL PAGE IS
OF POOR QUALITY

The repetitive scanning of a display in a man-machine system may become stereotyped if the scene/task appears frequently and requires a fixed level of performance on the part of the operator. For example, the task of flying an airplane using instruments for navigation requires skilled behavior, and dictates the presence of a relatively fixed scan pattern by the pilot (Weir and Klein, 1970; Waller and Flowers, 1977). Research on eye scanning of instruments in aircraft pilots dates from the work of Fitts and his associates (Jones, et.al., 1946). Indeed this work on probability of transitions between different instruments led to the regulations establishing the familiar "T" arrangement of the commonly used instruments in an aircraft cockpit:

AIRSPEED ATTITUDE ALTIMETER

DIRECTIONAL GYRO

Few other studies have been conducted on scanning behavior in pilots, probably owing to the complexity of instrumentation which has been required to perform such studies accurately. Several studies has strongly suggested the utility of scanning behavior in assessing a variety of human factors issues in the cockpit however. Dick (1980), for example has shown that there is a strong relationship between control inputs and visual scan strategy in pilots, demonstrating that there is typically a visual confirmation that a commanded input has achieved a desired change in one or more of the aircraft state variables. A recent study (Jones,et.al., 1982) also suggests the utility of using scanning information as an adjunct to pilot training. Both of these studies used the NASA/Langley oculometer to record eye scan. This device, based on the Honeywell oculometer, is suitable for conducting non-invasive scanning experiments in an aircraft cockpit (Spady, 1978). The work described here attempts to take advantage of this capability with an eye toward workload measurement techniques which may eventually be applicable during actual flight.

A CONCEPTUAL FRAMEWORK FOR THE STUDY

The results from some early experiments provided some insight into several flaws in the experimental design and the lack of basic knowledge of scanning behavior in general. Among the more salient problems identified were:

1. An unstated assumption of constant imposed mental loading throughout an experimental run was invalid since the piloting task requirements varied considerable in different segments of the approach. This problem is not uncommon however and exists in most of the previous pilot scanning studies. The Instrument Landing System (ILS) approach is often chosen as the piloting

ORIGINAL PAGE IS
OF POOR QUALITY

task in studies of workload (Waller, 1976; Krebs and Wingert, 1976; Spady, 1977). However, 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.

2. There was insufficient data in any segment of the run to allow a reasonable statistical analysis of scan factors. Since it was not known which factors, if any, in the scan were important, it was essential to first determine if any "steady state" effects were present in the eye movement patterns.

3. The levels of difficulty of the verbal loading task (see detailed description below) were not sufficient to induce large changes in the scanning pattern. Thus, while some trends were noted in the scan as a result of the additional imposed task, these were not consistent and at no time were any of the subjects even close to being heavily loaded.

4. There was not a range of pilot skill represented in the subjects; all were highly experienced and skilled NASA test pilots. It would seem very likely that inexperience pilots might perform rather differently in these types of experiments.

The above observations strongly suggested to the investigators that a more systematic, fundamental experiment might lead to more useful results. An inescapable conclusion may be drawn from these observations: Due to their interrelationships, workload, skill, and performance cannot be divorced from one another but must be studied together. The investigator must attempt to explicitedly control or at least have quantitative knowledge of each of these parameters in order to make sense out of any one of them.

As a guide toward experimental design and future data analysis, a conceptual model of pilot behavior was developed to aid in our thinking. It was felt that this model should include the following factors:

1. Performance - observed performance may be functionally related to all of the other factors; if the model is to be useful, it should predict situations in which performance will decrement

2. Pilot skill, including familiarity with the task(s) in a particular experiment. If he or she is unfamiliar with the task, learning may be expected during the course of an experiment

3. Inherent difficulty in the task(s) which are performed; some flight maneuvers are much more complicated than others

4. Nature and number of tasks which occur simultaneously with the primary task of flying the aircraft

ORIGINAL PAGE IS
OF POOR QUALITY

5. Psychological/physiological state of the pilot; probably quite important but not clear whether these are part of the independent or dependent variable

6. Random Noise

A hypothetical, graphical expression of there relationships is given in figure 1. Attempts at fitting a model using these parameters to the hypothetical situation in figure 1 will be presented later in this discussion.

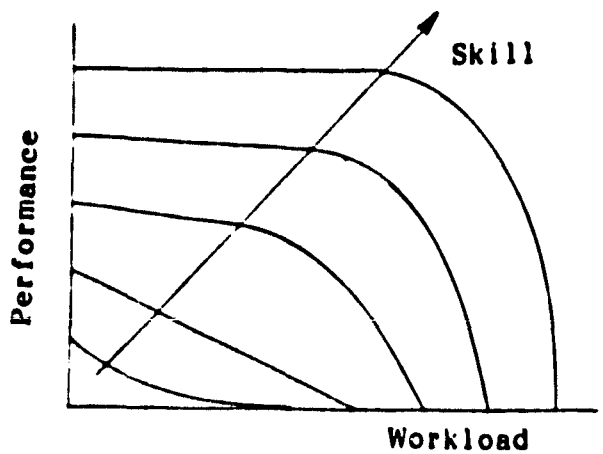

Figure 1. Hypothetical Relationship between Performance, Skill, and Workload

EXPERIMENTAL PROCEDURE

With these thoughts in mind, we set out to design a more straightforward series of experiments which would first consider whether it was possible to demonstrate consistent changes in the "steady state" scanning behavior during an instrument flight manuever of constant difficulty in the presence of some controlled variation in mental difficulty of an additional task. If it could be shown that the steady state behavior could be altered, one might then proceed to determine the shortest epoch over which a reasonable estimate of the effect might be made.

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

Piloting Task

As a piloting task, we chose a simple, yet realistic, steady state instrument manuever which might be expected to occur for periods of up to 10 minutes in actual flight. This time period was chosen as an estimate of the minimum amount of time required to provide a sufficient number of fixations to satisfy the assumption of steady state conditions. The task

ORIGINAL PAGE IS
OF POOR QUALITY

was to fly a precision straight and level course with zero degree glide slope and constant localizer sensitivity while maintaining a constant heading and airspeed in the presence of a low level of turbulence. A schematic representation of the task is presented in figure 2.

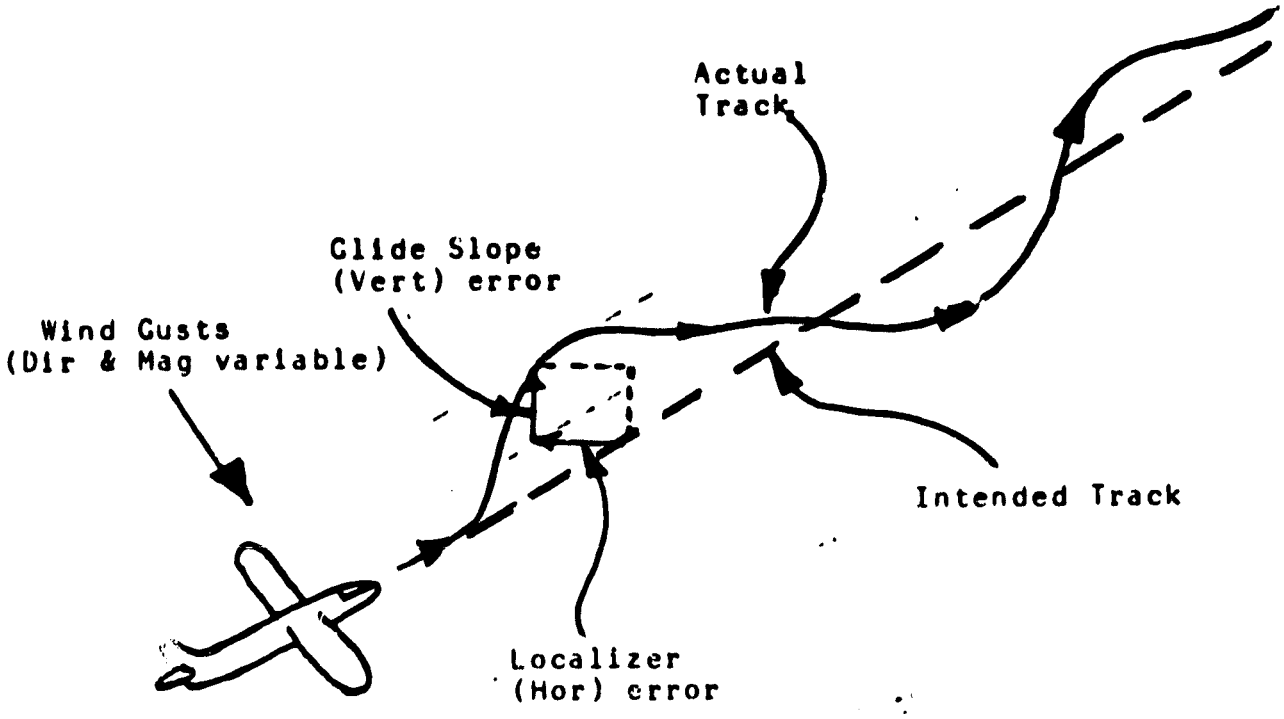

Figure 2. Schematic of Precision Straight and Level Flight

Pilot lookpoint on seven instruments (Attitude Indicator 'ATT', Directional Gyro 'DG', Altimeter 'ALT', Vertical Speed Indicator 'VSI', Airspeed 'AS', Turn and Bank 'T*B', and Glide Slope/Localizer 'GSL') was measured using the Langley oculometer. The oculometer can measure 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

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. This was accomplished by having the pilot respond verbally to a series of evenly spaced three-number sequences (Wittenborn, 1943). The pilot was told that he must respond to each three-number sequence by saying 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 =

ORIGINAL PAGE IS
OF POOR QUALITY

Positive Number Sequences:

Negative Number Sequences: All Others

Examples:	2 - 5 - 9	+
	8 - 3 - 6	+
	9 - 6 - 2	-
	3 - 7 - 4	-

Figure 3. Mental Loading Task Algorithm

"plus" (e.g. 1-2-3), otherwise, "minus" (e.g. 9-5-1). This algorithm is shown graphically in Figure 3. 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.

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

(1) $$TD = 1/\text{interval between task}$$

where TD is equal to imposed task difficulty.

In order to allow a wide range of loading, the task included 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 as calculated from equation (1). Calibration using the side task described below confirmed the relative difficulty of these number intervals.

Numbers were generated by a computer controlled speech synthesizer (see hardware description below). 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. Performance was recorded by having the pilot press a 3-position rocker switch mounted on the yoke up for plus and down for minus.

Visual Side Task for Workload Calibration

The amount of mental loading imposed on the pilot by the number task was calibrated using a side task. The runs made with

ORIGINAL PAGE IS
OF POOR QUALITY

the side task were not used in the scanning analysis, however, due to the alteration of normal scanning caused by the task. The side task employed a CRT which could display an asterisk appearing in the upper half or in the lower half of the screen. The display was mounted to the left of the simulator just outside the pilot's peripheral view. The asterisk appeared at random intervals between one and three seconds and remained on for one second (Ephrath, 1975). The pilot was told to turn the symbols off by using a three position rocker switch on the control grip. Moving the switch upward turned the upper asterisk off, downward turned the lower asterisk off. This task was done only when the pilot had time left from performing the primary tasks of flying the airplane and answering the number task. Thus the number of correct responses on the side task gave a measure of the residual capacity of the pilot from which a workload index could be calculated. The expression used to calculate the workload is given below. The constants were obtained using the best least squares fit weighting coefficients.

$$\text{(2)} \qquad WLX = \frac{(.780)(RT) + (.626)(MISS)}{(.780 + .626)(NSTIM)} \times 100 \text{ percent}$$

where:

WLX = workload index
RT = cumulative response time (seconds)
MISS = number of incorrect responses
NSTIM = total number of stimuli (symbols) presented

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. Eleven subjects ranging in skill from NASA test pilots to non-pilots participated in the experiments.

EQUIPMENT

A desktop general aviation instrument flight simulator (Analog Training Computers ATC-510) was used to simulate the piloting task. 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.

ORIGINAL PAGE IS
OF POOR QUALITY

The NASA/Langley Oculometer is described elsewhere (Middleton, et.al., 1977; Spady, 1978) and the interested reader is referred to these documents. For the experiments described here, the oculometer provided a discrete voltage level corresponding to the current instrument fixation. This level was based on pilot lookpoint falling within predetermined X-Y boundaries about each instrument on the simulator panel.

The simulator panel and oculometer optical head are shown in figure 4.

A general purpose 8085 microprocessor development system (Burns, et.al., 1979) was used to control the verbal task and the workload calibration side task as well as to digitize, store, analyze, and display the scanning data from the experiments described here. The system was equipped with 64K of RAM, an 8085 processor, two serial ports, an 8 channel/12 bit A/D converter, a CRT controller, a speech synthesis module, two double sided double density floppy disk drives with a Shugart 1403D intelligent controller module, and a dot matrix graphics printer. A photograph of this system is shown in figure 5. Software for the system was written in STCIC, an interactive programming language based on FORTH (Sachs, 1980) and in 8085 assembly language. Details of the programs may be found in the thesis by Stephens (1981).

Aircraft performance data was recorded during each of the experimental runs. The data recorded included : x-coordinate of lookpoint, y-coordinate of lookpoint, track/no track, pupil diameter, instrument identification number, glide slope indicator deflection, localizer indicator deflection, elevator deflection, aileron deflection, pitch attitude, and roll attitude. These signals were recorded on a 14-channel FM tape recorder, and digitized at NASA/Langley. Later the digital representations were transferred to floppy disks on the microprocessor system. The RMS error and frequency content of the glide slope and localizer indicator deflections were used to define the aircraft performance for each run (see later discussion).

INDEPENDENT ESTIMATE OF PILOT SKILL

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.

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

ORIGINAL PAGE IS
OF POOR QUALITY

LABORATORY EQUIPMENT

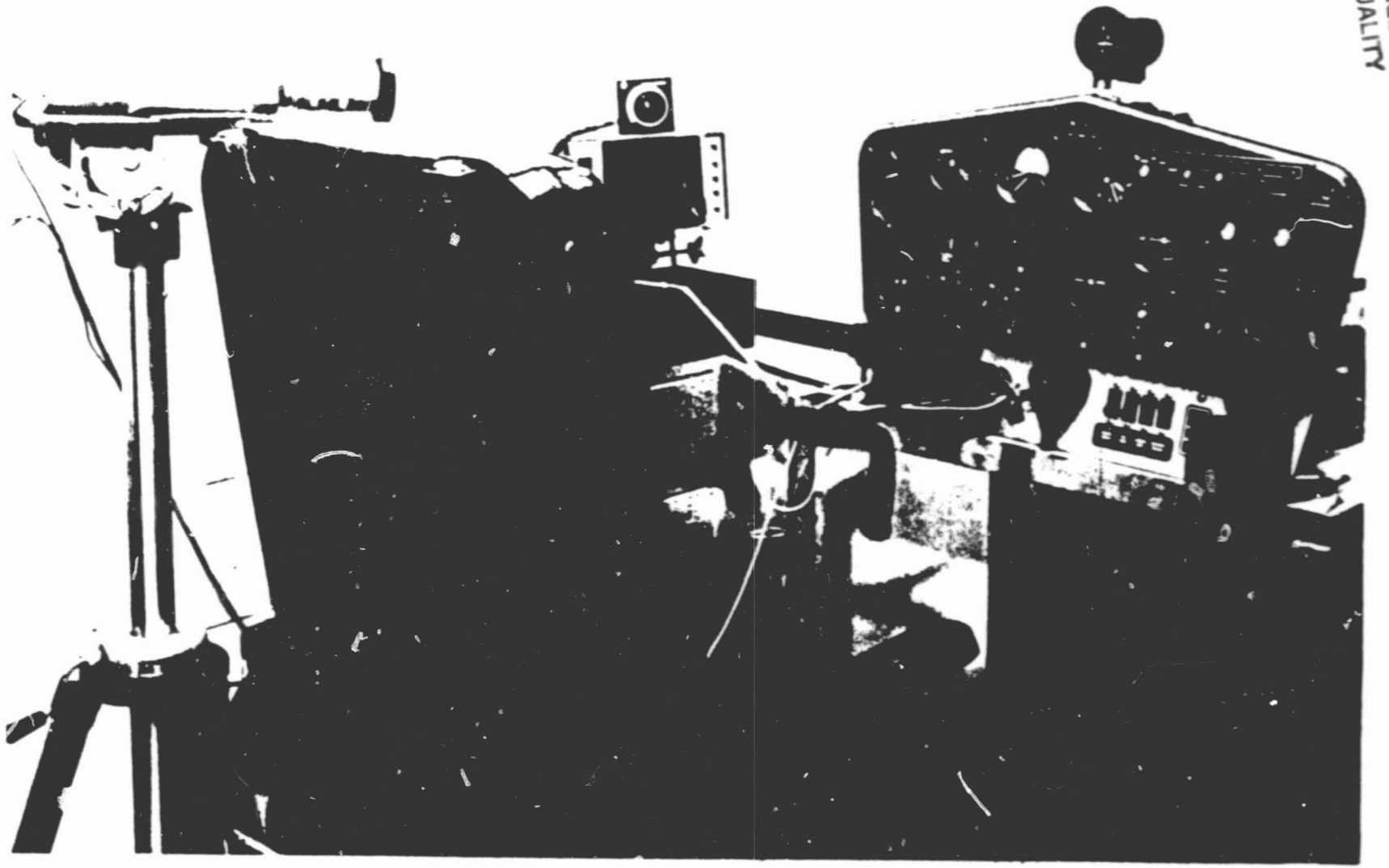

Figure 4. ATC 510 Simulator with Oculometer Head

ORIGINAL PAGE IS
OF POOR QUALITY

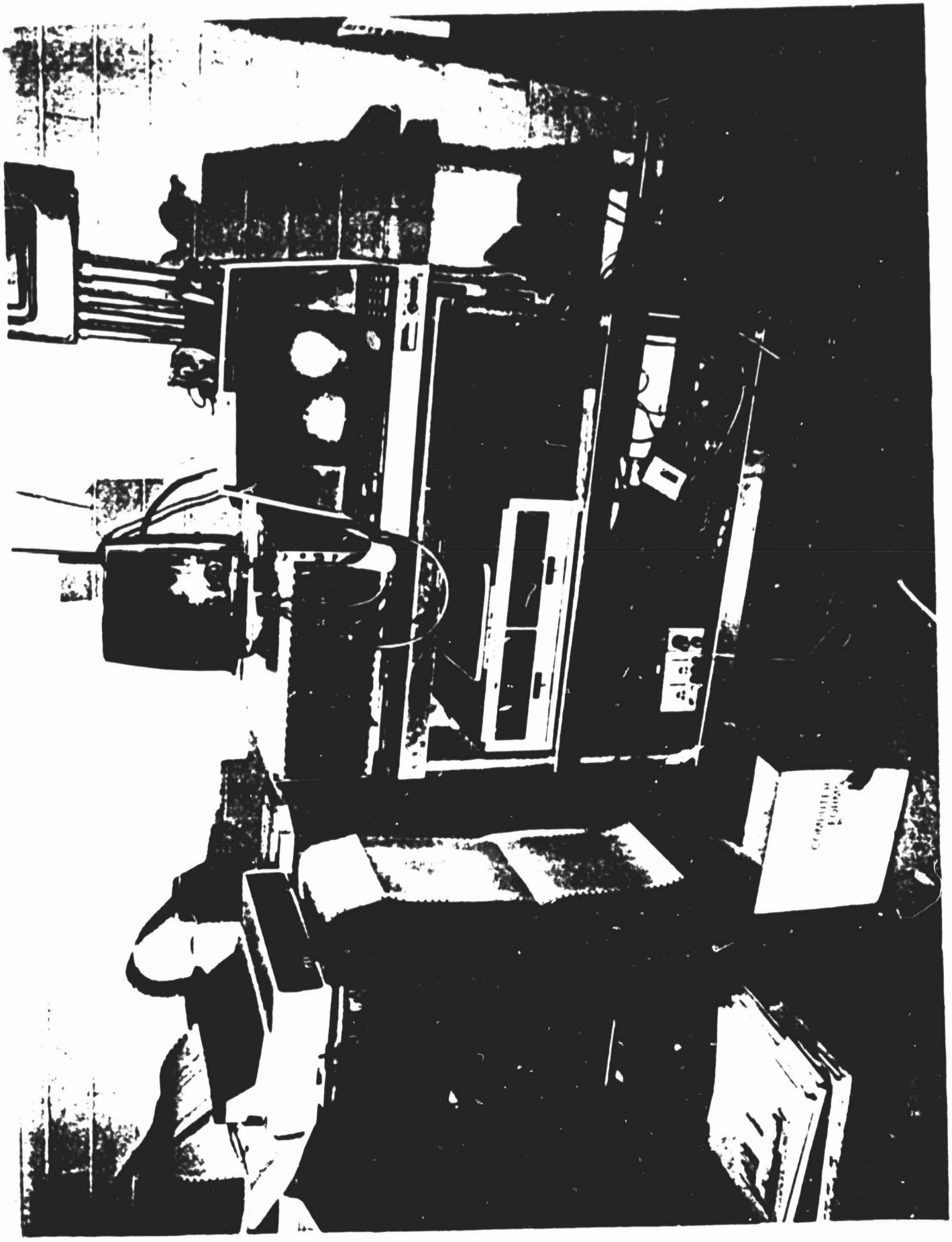

Fig 5.Computer System used for Experimental Control and Data Analysis

ORIGINAL PAGE IS
OF POOR QUALITY

where

Skill = score reflecting relative piloting performance
recency = number of flight hours in past 30 days
totaltime = 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 or rating
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 Number	Skill Score(%)
3	100.00
4	85.31
11	76.64
13	53.96
15	38.81
6	37.47
12	33.23
14	31.71
8	22.74
7	15.28
16	12.83

Table I. Relative Skill Scores of all Subjects

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.

RESULTS

Initial Data Analysis

A set of preliminary experiments using this protocol and apparatus were conducted during the summer of 1980. Subjects with a wide range of skills, from non-pilots to NASA test pilots, participated.

ORIGINAL PAGE IS
OF POOR QUALITY

Ten minute runs with the side task were performed with 3 of the pilots. The workload index defined above were determined for each pilot for all loading levels (Table II). The index increased monotonically for all subjects with increased rate of presentation of the number task. The average workload index varied from 80 percent for no mental loading task to 92 percent at the 4 second interval and 96 percent at the 2 sec intervals. Although we were not able to evaluate the workload index with all pilots, the results with these three pilots did allow us to confirm quantitatively that the mental loading is increased as the interval between number presentations decreases.

Pilot Number	No Loading	4-sec Intervals	2-sec Intervals
9	87	93	95
5	82	94	97
7	70	89	--
Average	80	92	96

Table II. Workload Sidetask Results

Dwell Time Histograms

The raw scanpath data is of the form lookpoint vs. time. An example of the raw data is shown in figure 6. From this data dwell time histograms may be plotted for each instrument in the scanpath. Examples of the results from several of these experiments are shown in Figure7.

In the four novice subjects, the dwell time on the primary instrument (the Attitude Indicator in all but the non-pilot who used Glide Slope/ Localizer) became progressively weighted toward extremely long dwells as the verbal task difficulty increased. Figure 7 shows the dwell time histograms for all pilots on the Attitude Indicator, Directional Gyro, Glide Slope/Localizer and Vertical Speed Indicator. First consider the plots for subject #5 who has intermediate skills. Note that for the no loading case, the dwell histogram on the Attitude Indicator of subjects #5, #9 and #10 has a fairly standard shape (Harris and Christhilf, 1980). When numbers are added to the piloting task, the dwell becomes longer and the mode of the histogram at 1/2 second begins to disappear. The effect is even more dramatic for 2-second interval case; the entire distribution is skewed toward extremely long dwells on Attitude as the pilot apparently begins to "stare" more and more at this instrument. Similar effects are seen for pilots 9 and 10.

An interesting difference occurs for subject # 7, the non-pilot, however. This subject had no previous piloting

ORIGINAL PAGE IS
OF POOR QUALITY

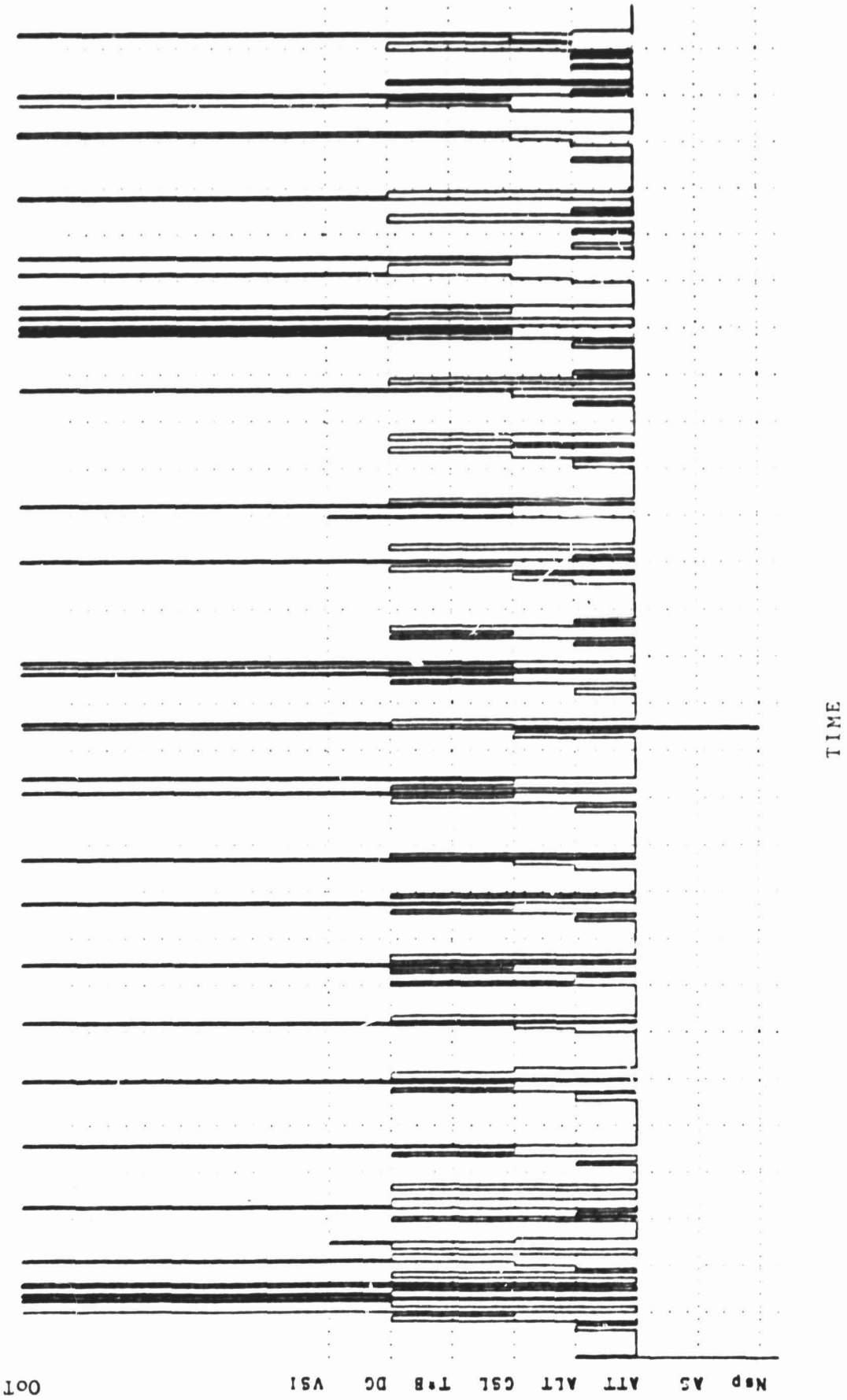

Figure 6. Raw Scanning Data

ORIGINAL PAGE IS
OF POOR QUALITY

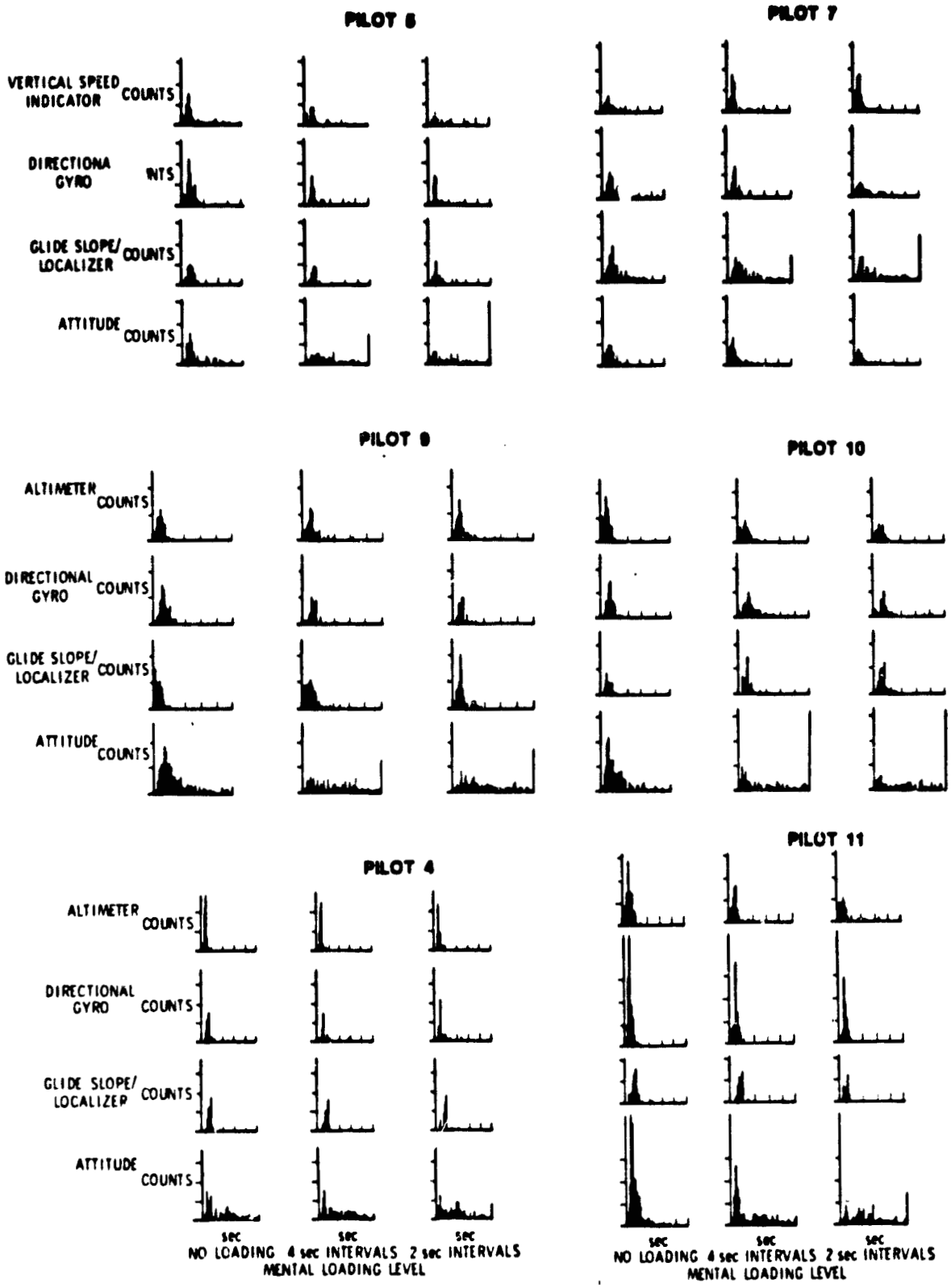

Figure 7. Dwell Time Histograms

ORIGINAL PAGE IS
OF POOR QUALITY

experience and was only given enough practice to allow him to stay nominally on course during the precision straight and level maneuver. Note that this subject adopted the Glide Slope/Localizer as the primary instrument apparently in an effort to accomplish the precision task by keeping the needles centered. Even though the subject adopts the inappropriate instrument to accomplish the piloting task, the dwells on this instrument are affected in a manner similar to those on Attitude for the more experienced subjects.

The visual scanning behavior of the two subjects with higher levels of skill was also affected by the verbal loading (subjects 4 & 11 in Figure 7). However, the effect was much less than seen in the novice pilots. Figure 7 also shows the dwell time histograms for the NASA test pilot, subject #4. Note that he develops a slight stare on the Attitude Indicator for the highest loading condition but his histograms are otherwise unaffected. Subject #11, who had the next highest skill level, was somewhat more affected, especially at the highest loading level, as indicated by the histograms for the Attitude Indicator (Figure 7). Subject #11 uses a large number of short dwells on the Attitude Indicator under the no loading case. When the mental loading task is introduced at 4-second intervals, his distribution is shifted to somewhat longer dwells. However, there is still a very significant peak at around 1/2 second. The actual shift in dwell times is not as large as that seen in the novice pilot's histograms, even though there appears to be a large change due to the reduction in magnitude of the histogram peak.

The shift to longer dwells may also be demonstrated by looking at the percentage change from the no loading case in the number of dwells on the primary instrument that are 5 seconds or longer in duration as the mental workload is changed. The raw counts of such dwells are shown as the last element in the histograms. Table III shows the percentage change from the no loading case for each pilot. The percentage of dwells is seen to increase with decreasing skill level. This holds for all subjects except subject #7, the non-pilot. It should be pointed out, however, that subject #7 used a different primary instrument from the rest of the pilots and therefore had a completely different basic scan pattern from the other pilots. This fact may not allow direct comparison of the results from subject #7 with the other subjects. This is not a cause for concern since the results from all of the pilot subjects seem to be consistent and, therefore, any conclusions drawn from their results should be applicable to other pilots.

The dwell time characteristics of secondary instruments were most affected in the novice subjects. The secondary instrument dwells are seen to change in a different manner than the primary instrument dwells. As opposed to the shift to longer dwells, as in the case for the primary instruments, the effect of loading in the secondary instruments is to decrease the number of looks at that instrument, perhaps an example of a

ORIGINAL PAGE IS
OF POOR QUALITY

phenomenon known as load shedding. The shape of some of the histograms changes under varying loading conditions. Subject #4 was the only subject whose dwell time histograms on secondary instruments were not affected by loading. Subject #11 appears to exhibit some load shedding, primarily on the Altimeter and Vertical Speed Indicator.

Pilot Number	No Loading	4-sec Intervals	2-sec Intervals
4	0	0.6	3.7
11	0	1.95	7.33
9	0	6.80	8.46
5	0	8.59	20.08
10	0	19.80	23.39
7	0	6.90	13.21

Table III. Percent of Primary Instrument Dwells Greater Than 5 Seconds

Fixation Sequences

It was of interest to examine whether pilots develop a scan pattern or patterns during the constant flying maneuver in this experimental paradigm. If the dwell times on individual instruments are ignored, an ordered list of instrument fixations may be developed for each pilot for the various loading cases. These lists may be broken up into smaller segments (or sequences) of various lengths for easier analysis. Each different sequence may be considered as a component of the overall scan pattern. One may hypothesize that those sequences which occur most frequently during the maneuver are those of most importance to the pilot and ones which might indicate an ordered scan pattern.

Examination of the results indicated that sequences of four-instrument fixations were the longest for which there was a significant amount of repetition during a run, hence sequences of length four were chosen for analysis. The number of times each four-instrument sequence occurred during a ten minute run was obtained as was the total number of sequences of length four in the run. From these data, the percentage of occurrence was calculated for each observed sequence. For example there might be 800 sequences of length four in 10 minutes. If the sequence, ATT-DG-ALT-DG, occurs 40 times during the run, its percentage of occurrence would be 40/800 X 100 percent = 5 percent. In this fashion, the percentage of occurrence of all length-four sequences in the no-loading case was determined for each pilot. The 10 sequences which occurred most frequently for each pilot were arbitrarily chosen as indicators of the scan patterns normally used by the various pilots. In general, the specific sequences were different for each pilot. The manner in which the percentage

ORIGINAL PAGE IS
OF POOR QUALITY

occurrences for these 10 sequences change for each subject as a function of mental loading is shown in figures 8. Figure 9 plots the sum of these percentages across loading for all the subjects. It is important to note that the sequences used as the basis for calculation for all conditions are the 10 most frequent for the no-loading case. Each line beginning at the no loading case and ending at the 2-sec interval case represents the same sequence.

Several interesting observations may be made by comparing the plots of the skilled pilots (figure 8e and f) with those of the novice subjects (figures 8a-d). A difference may be seen between the two groups in the percentage of occurrence of the most often used sequences. The first 10 sequences used by the skilled pilots comprise over 50 percent of their scan pattern (see sum in figure 8). The usage of these 10 sequences is relatively constant with changes in loading suggesting that the patterns are not disturbed by the verbal number task. The novice pilots' results differ in several respects from those of the skilled subjects however. The 10 most frequently used sequences in the no loading run occupy much smaller percentages of the total scan than do those of the skilled pilots. This suggests the novices' scans are more random than those of the skilled subjects, even without the imposition of an additional task.

The novice subjects also show a consistent decrease in the percentage occurrence of the 10 sequences as the workload is increased. This decrease may be the result of either the equalization of the number of occurrences of each sequence in the run (i.e. a trend to randomization) or a change to a different set of sequences from those used in the no loading case.

These findings both strongly supported the possible utility of the instrument scan as an indicator of both workload and skill. However, neither method seemed to allow direct comparison between scanpaths for different types of maneuvers since instrument usage might vary considerably for different tasks. It thus appeared important to develop a more general analysis method.

Quantifying Disorder in the Scanpath

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 a general method for the study of scanning behavior. To be most useful the method should be independent of the number and arrangment of

ORIGINAL PAGE IS
OF POOR QUALITY

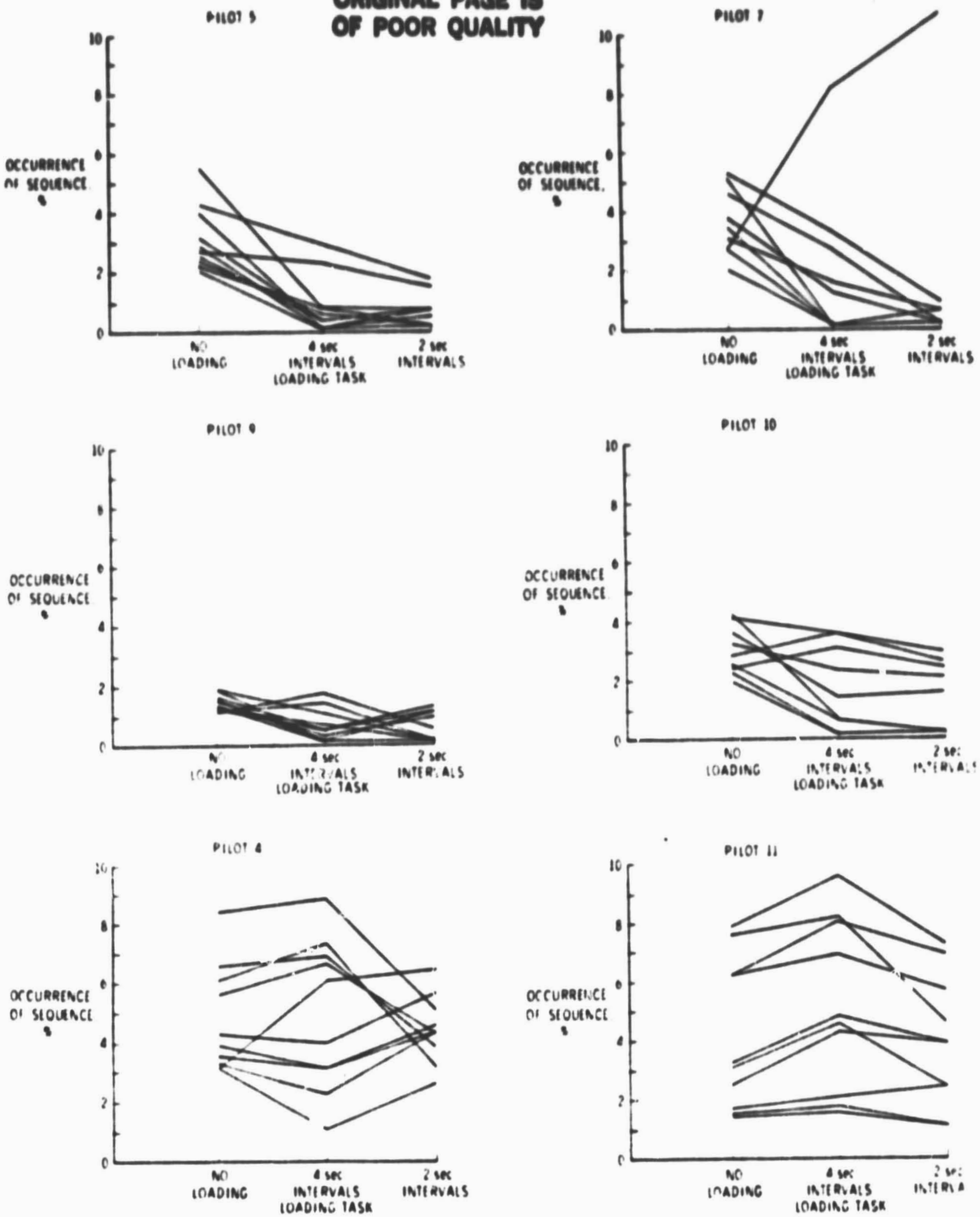

Figure 8. Percent Occurrence of Sequence vs Loading Task

ORIGINAL PAGE IS
OF POOR QUALITY

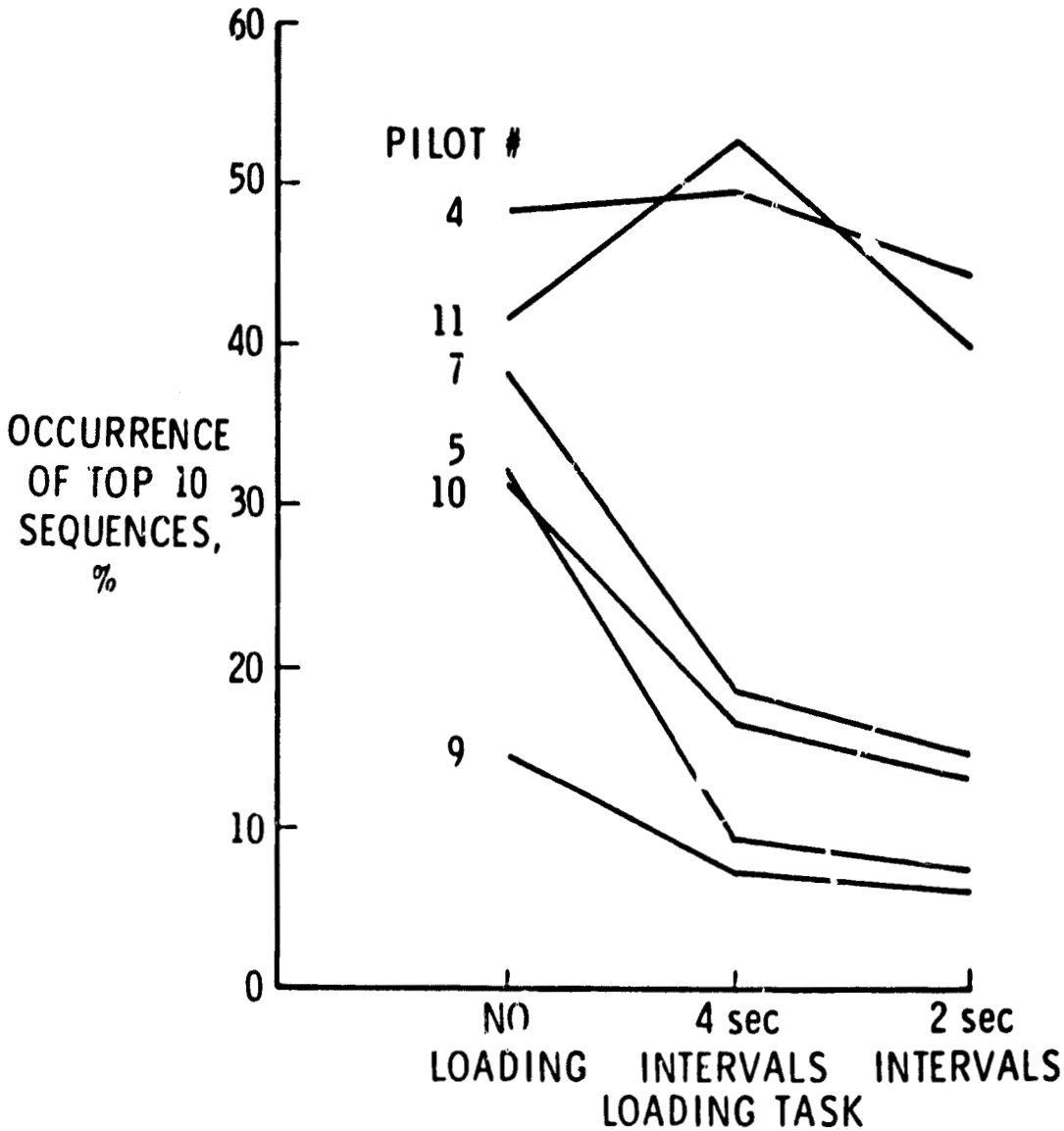

Figure 9. Percent Occurrence of Sequence vs Loading Task

ORIGINAL PAGE IS
OF POOR QUALITY

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.

The piloting task in the current experiment is such that the pilot's scan can only lie either on one of the seven specified instruments or on outside the oculometer's range. Each fixation may be of arbitrary duration. The time history of fixations has a form which is similar to that of a communication system which can assume eight discrete states with a varying duration in each state (see figure 6). The orderliness of such a system is related to the probabilities with which it occupies 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):

$$H_o = - \sum_{i=1}^{D} [p_i \log_2 p_i] \tag{4}$$

where H_o = observed average entropy
p_i = probability of sequence i occurring
D = Number 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.

Note: The term Entropy has been associated with Information Theory for so long that its usage tends to suggest an attempt to quantify the information content of some system. However, older usage of the term comes from thermodynamics where entropy is used to describe the amount of disorder present in a system. In the present discussion it must be emphasized that there is no attempt to quantify the amount of information which the pilot is acquiring from his or her displays. Rather the mathematical form for entropy is used to compactly describe the amount of spatial and/or temporal order present in the pilot's scanpath, in keeping with the meaning of entropy in thermodynamics.

ORIGINAL PAGE IS
OF POOR QUALITY

In order to calculate the entropy of the scan, each of the instruments to be examined was given a number. As the pilot scanned the instrument panel a sequence of these numbers was then stored 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 since these seemed to yield the most consistent results. The remainder of the discussion here applies to the results for length 2 sequences. Details of the methodology are given elsewhere (Stephens, 1981).

Note: For short observation times, 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 4 below) which occured 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, H_{max}. H_{max} 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:

(5) $$Q = M \cdot (M-1)^{N-1}$$

= maximum number of sequences of length N

or

The number of bits required to uniquely encode all Q possible sequences is $\log_2 Q$. It represents H_{max} of the visual scan for the number of instruments and sequence length being considered. For example, with 8 states (7 instruments + out of range) the value of Q for sequences of two instruments is 56 which yields a corresponding $H_{max} = 5.8$.

The normalized value of H may then be calculated from:

(6) $$H_{corr} = H_o \cdot A/H_{max}$$

where $A = \mathrm{Log}_2 L$ for $L<Q$; = 1 otherwise
$L = R-N+1$ = number of sequences in a run
R = number of fixations in a run
N = sequence length (N = 1,2,3, or 4)

A Revised Method for Calculating Entropy

The method for calculation of entropy described above has a flaw which had to be corrected in order to insure proper calculation of frequency of occurence of different sequences. The method described above ignores the overlap between

ORIGINAL PAGE IS
OF POOR QUALITY

successive sequences. For example, the sequence 123431431 is interpreted to include the length four sequences 1234 2343 3431 4314 and 1431. Clearly, the frequency of sequences determined in this fashion will be correlated and in fact does not provide the appropriate estimate of probability of sequence occurence. Consider the sequence 12121212. For purposes of our analysis, it probably does not matter whether the sequence 1212 or 2121 is considered to occur. Both relate essentially the same pattern when a long run such as this occurs. The pattern 12125342121 on the other hand shows these sequences to be different on the basis of context in the scan pattern.

Recognizing this problem, we have adopted a new method of calculating the frequencies of the various sequences. An initial pass is made on the data using the original method to identify sequences. That sequence which occurs most frequently is noted, the number of occurences stored, and the occurences of this sequence are then removed from the data run by inserting -1 instrument code in the relevant locations. A second pass is then made in which the most frequent valid sequence (the -1 codes are ignored) is identified and removed. This process continues until all independent sequences have been identified and removed. This process insures that no sequence is counted twice in estimating the probabilities of occurence of different sequences.

Entropy Rate

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 of 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:

$$H_{rate} = H_o/t \tag{7}$$

where H_o is the entropy for the system given by equation 2 and t = smallest interval in which that transition occurs.

In practice, H_{rate} is an average value given by the following:

$$H_{rate} = \sum_{i=1}^{D}[(H_{corr})_i/DT_i] \tag{8}$$

where $(H_{corr})_i$ = Normalized entropy for ith sequence
DT_i = Average dwell time for ith sequence
D = Number of different fixation sequences

The maximum value which H_{rate} can assume may be calculated using the H_{max} determined above together with dwell time

ORIGINAL PAGE IS
OF POOR QUALITY

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 (Fahill, 1977). A more realistic average value is probably about two fixations/sec or less for a long period of instrument scan (say > 10 sec).

Using this value (0.5 sec/look) as the average dwell interval, the maximum entropy rate for sequences of length two is calculated from equation 5 to be:

$$(H_{rate})_{max} = 5.8/0.5 \;.\; 2 \text{ fixations/seq.} \cong 6 \text{ bits/sec}$$

This number represents an upper bound. Since we suspect that the pilot must exhibit 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 H_{rate} 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 H_{rate} tend toward lower values under such conditions. Figure 10 plots H_{rate} vs number task difficulty for several pilots.

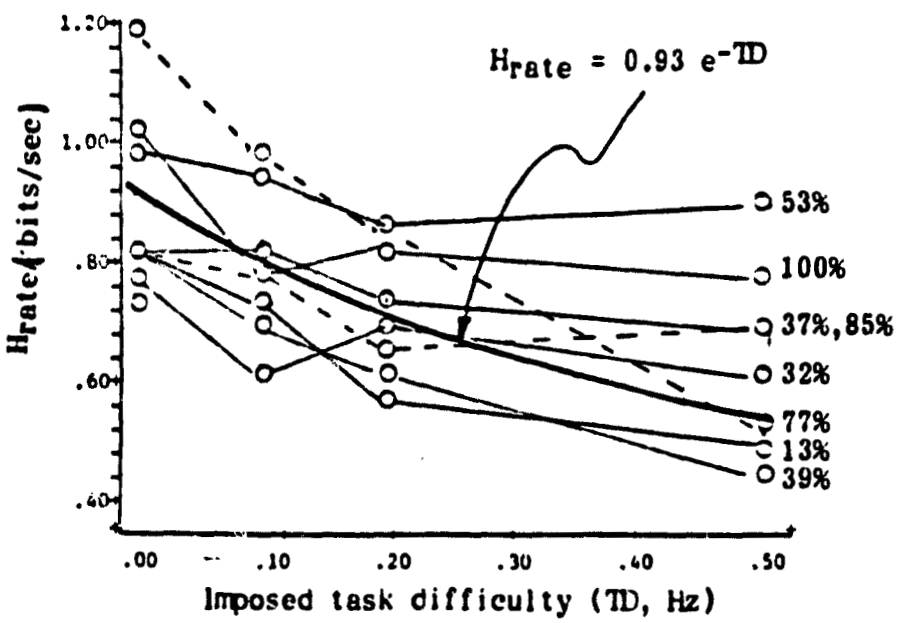

Figure 10. Entropy Rate on Length-2 Sequences vs Imposed Task Difficulty for 8 Pilots (Relative Skill Levels Shown on the right - highest=100%)

ORIGINAL PAGE IS
OF POOR QUALITY

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 H_{rate} at all loading levels ($p < 0.01$) and equality of mean H_{rate} between subjects ($p < 0.01$). All six combinations of level differences in mean H_{rate} were found to be statistically significant (T-test $p < 0.05$). Thus H_{rate} was chosen to map from scanning behavior into task difficulty (i.e. workload).

The model used expresses H_{rate} as an exponential function of TD.

(9) $$H_{rate} = 0.9279\ e^{-TD}$$

This equation was obtained via a regression analysis based on the data from seven of the pilots with a coefficient of determination, $R^2 = 97.3\%$. It is solved for task difficulty with the following result:

(10) $$TD = -[0.06 + \ln(H_{rate})]$$

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

Autocorrelation and Power-Spectral Density

Another analysis method is the autocorrelation of the instrument scan pattern. The purpose of this particular method of analysis is to determine whether or not the pilot's scan is altered by the mental loading number task in a periodic fashion. One possible alteration that might be encountered is that the frequency at which an instrument is sampled may change as the auditory task changes. Specifically, the nature of the relationship between instrument scan frequency and number task presentation frequency task would provide valuable hints on how the task, and therefore the associated mental load, affects the scanning pattern.

The autocorrelation was performed on the data as described below. Due to the arbitrary nature of the assignment of instrument numbers, the autocorrelation of the signal containing all instrument numbers would not necessarily produce meaningful results. For this reason each of the seven instruments were examined successively by replacing the time sequence of all instruments with a sequence $\{x_j(i)\}$ where the value is 1 when instrument j is being fixated and 0 when any other instrument is being fixated. In order to eliminate the dc component for further spectrum analysis, a zero-mean sequence $\{f_j(i)\}$ was computed from $\{x_j(i)\}$ as follows:

(11) $$f_j(i) = x_j(i) - \bar{x}_j$$

ORIGINAL PAGE IS
OF POOR QUALITY

where $x_j(i) = 1$ if specified instrument j is being fixated and 0 otherwise
$\bar{x}_j$ = mean of $\{x_j(i)\}$

The sample autocorrelation of $\{f_j(i)\}$, or sample autocovariance of $\{x_j(i)\}$, was calculated by the formula:

$$R_j(k) = 1/n \sum_{i=1}^{n} [f_j(i) \cdot f_j(i+k)] \tag{12}$$

where $R_j(k)$ = autocorrelation sequence for instrument j
n = number of samples = total run duration/oculometer sampling period (1/30th sec)

This autocorrelation was computed for each of the seven instruments for each loading case on each pilot. In order to detect possible periodicity in the scan, the Fourier transform of the autocorrelation was taken to produce the power density spectrum. From this a value for the dominant frequency may be obtained.

The power-spectral density was obtained by using a Fast Fourier Transform (FFT) package available on the microprocessor system. Some interesting results emerged from this analysis the first of which may be seen in Figure11. This shows the autocorrelations for pilot #4 (second highest skill level) for his attitude indicator on each of the four different mental loading cases. A change in the dominant frequency may be seen as the loading is increased. The power-spectral densities shown in Figure12 show the dominant frequencies for the low (10-second intervals), medium (5-second intervals), and high (2-second intervals) levels of mental workload to be 0.0928 Hz, 0.1709 Hz, and 0.3175 Hz respectively. These frequencies correspond to periods of 10.78 seconds for the low, 5.84 seconds for the medium, and 3.15 seconds for the high level of mental workload. These periods are closely related to the number tasks periods (11, 6, and 3 sec) given by the sum of the interval between number presentation and the time required to present the numbers. This implies, at least for this pilot, that the loading task directly influences the scan pattern. When no numbers are presented, the pilot scans his instruments in a close-to-random manner and the density spectrum exhibits no dominant frequency (cf fig.12.a). When the periodic task is applied, the scan becomes more and more periodic with increased task frequency (cf fig.12.b&c). This demonstrates that the pilot has a tendency to multiplex the flying task and the number task for greater efficiency. Overload occurs when numbers are presented too rapidly for the pilot to efficiently multiplex both tasks (cf fig.11.d). A similar behavior is observed for all of the higher skilled pilots as demonstrated in Table IV. The periods of oscillation for the 5 pilots of highest skill appear to match those presented to them by the mental loading task very closely. However, the other 6 pilots do not seem to have any consistent pattern in their autocorrelation of sequences. Most of the

ORIGINAL PAGE IS
OF POOR QUALITY

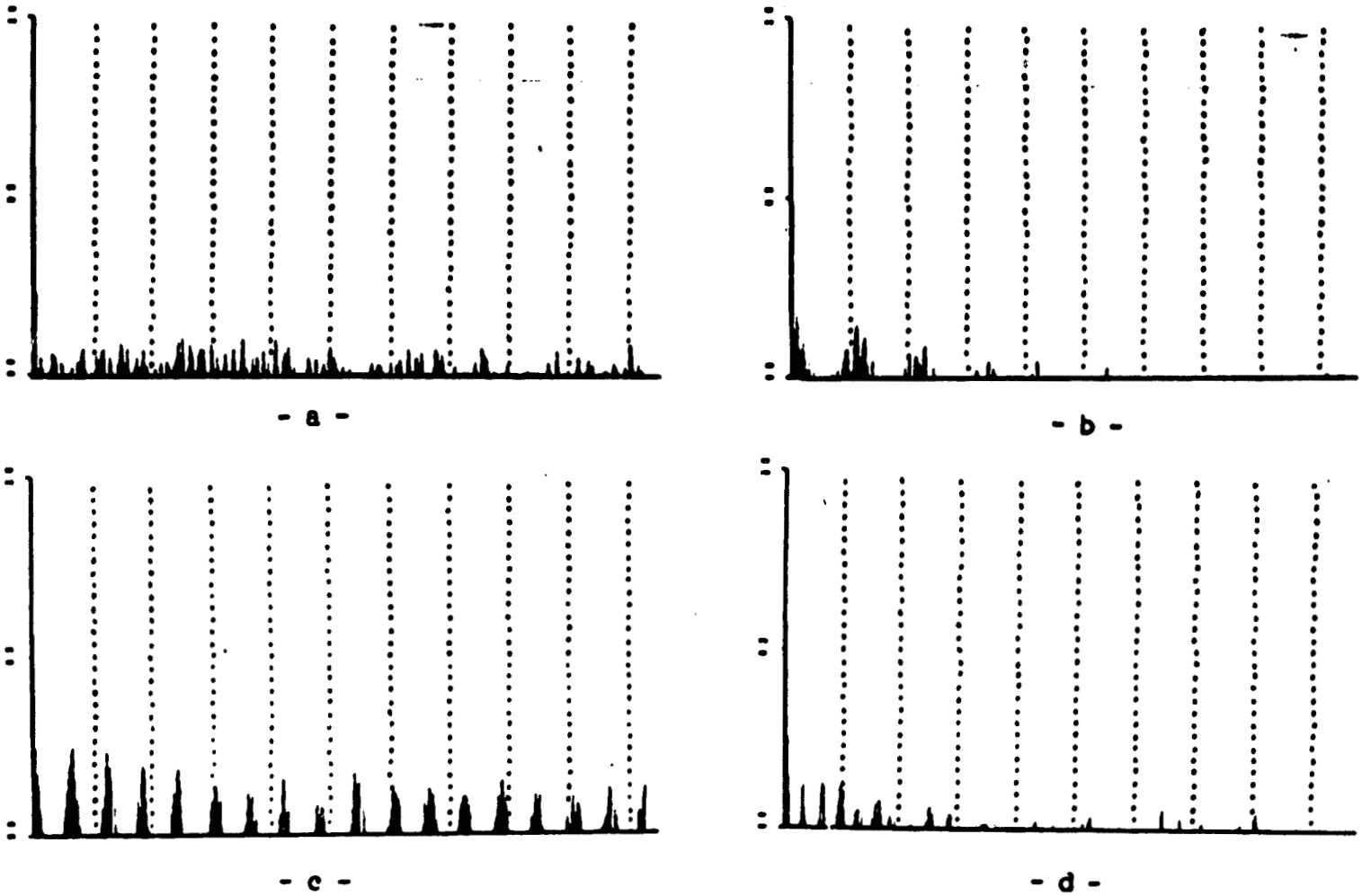

Figure 11. Autocorrelation for Pilot #4 (relative skill levels = 85%) using Attitude Indicator (Dotted Lines Indicate 10-sec Intervals). Number Task Intervals and Associated Task Difficulties are a) No Intervals - 0, b) 10 sec - 0.1, c) 5 sec - 0.2, d) 2 sec - 0.5

ORIGINAL PAGE IS
OF POOR QUALITY

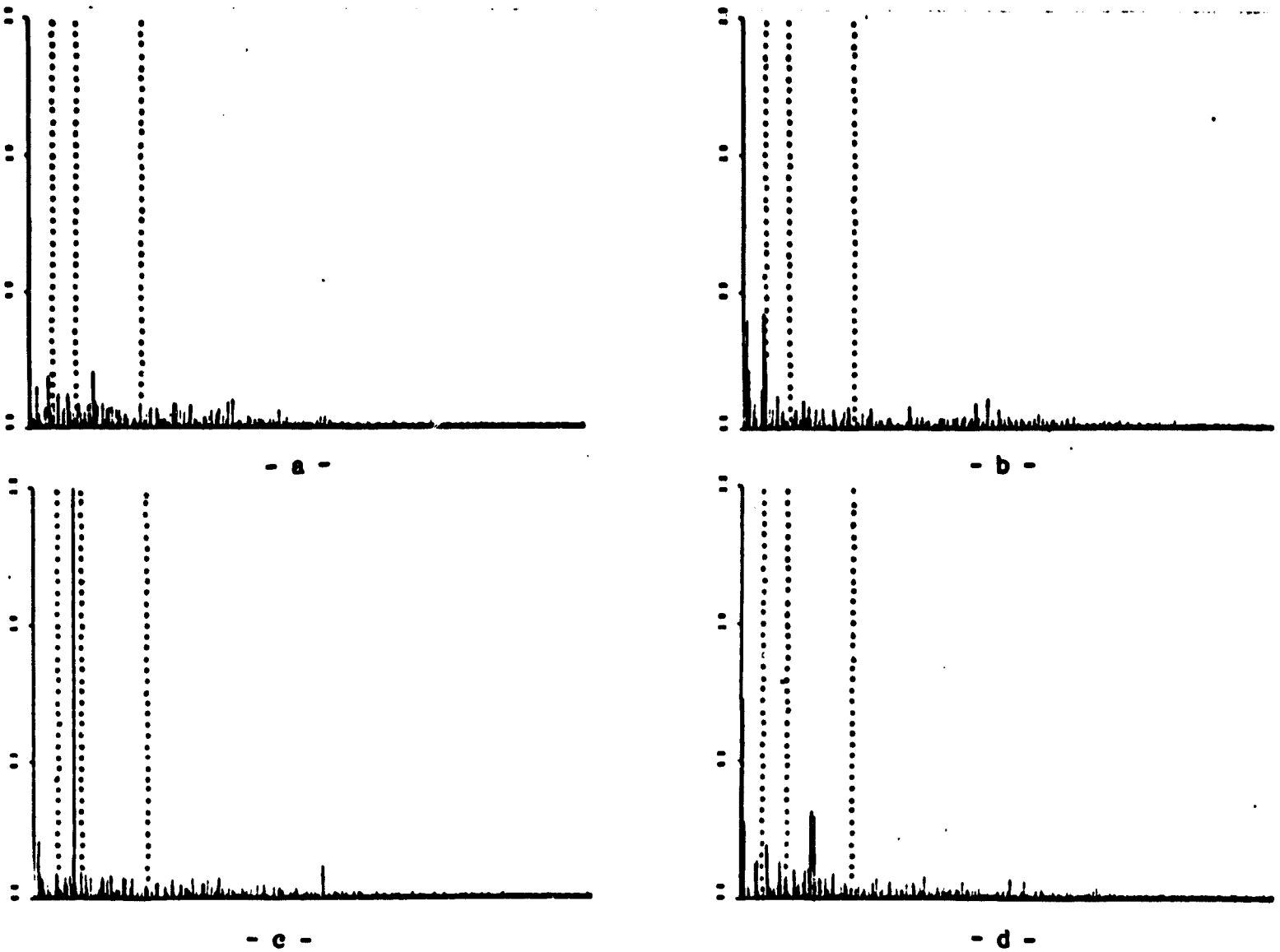

Figure 12. Power Spectral Densities for Pilot #4 (Relative Skill Level = 85%) Using Attitude Indicator (Dotted Lines correspond to Frequencies of 0.1, 0.2, and 0.5 Hz respectively). Number Task Intervals and Associated Task Difficulties are a) No Intervals - 0, b) 10 sec - 0.1, c) 5 sec - 0.2, d) 2 sec - 0.5

ORIGINAL PAGE IS
OF POOR QUALITY

		Number task periods(sec)		
		11	6	3
Pilot's relative skill levels	100%	9.75	5.69	4.18
	85%	10.78	5.85	3.15
	77%	9.75	6.40	6.02
	*53%	9.31	5.25	2.84
	39%	9.75	6.40	2.93
	37%	10.24	5.25	34.13
	33%	2.03	7.59	12.80
	32%	5.25	5.69	6.61
	22%	9.31	12.80	3.79
	15%	1.32	7.88	13.65
	13%	17.07	20.48	7.88

Table IV. Scan autocorrelation dominant periods for 9 pilots using attitude indicator (glide slope/ localizer for *) for 3 frequencies of the mental loading task.

pilots showed little or no periodicity in the no-loading case. One possible explanation of these results may be that the higher skilled pilots adapted their scanning to the task much faster and better than the lower skilled subjects. DeMaio, et al (1970) found that skilled pilots evidently developed optimum scanning strategies when presented novel tasks much faster than unskilled pilots. Another explanation may be that skilled pilots have a better developed ability to time multiplex several simultaneous tasks.

Performance Measures

Before discussing the modelling effort in this study, it is necessary to mention how task performance was estimated in these experiments. Several variables were obtained from each of the two tasks 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

ORIGINAL PAGE IS
OF POOR QUALITY

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.

$$(13) \qquad TP = \frac{(TOT - WRO - MIS)}{TOT} \times 100\%$$

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.

ORIGINAL PAGE IS
OF POOR QUALITY

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:

$$P = CONST + a(TP) + b(RMS/GS) + c(RMS/LOC) + d(\%PWR/GS) + e(\%PWR/LOC) \quad (14)$$

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 than 0.1 Hertz
%PWR/LOC = percent of power from the power-spectral density for the localizer greater than 0.1 Hertz

A Model Relating Workload, Performance, and Skill

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 the other two parameters. A model relating these three parameters 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:

$$P = P(0) - EXP((TD/Skill)^2) \quad (15)$$

This equation may be rearranged as follows:

$$EXP((TD/Skill)^2) = P(0) - P \quad (16)$$

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. Making use of equation (14), the right hand side of (16) may be expanded to yield:

ORIGINAL PAGE IS
OF POOR QUALITY

$$(17)\quad P(0) - P = a(\#TP(0) - \#TP) + b(RMS/GS(0) - RMS/GS) + c(RMS/LOC(0) - RMS/LOC) + d(\%PWR/GS(0) - \%PWR/GS) + e(\%PWR/LOC(0) - \%PWR/LOC)$$

A multiple regression analysis was then performed the expanded version of equation 16 using values for each of the indicate parameters 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.

The results of the first attempt at regression 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 17 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:

$$(18)\quad EXP((TD/Skill)^2) = 1.4483 + 0.0351(\#TP(0) - \#TP) + 0.1765(RMS/GS(0) - RMS/GS) - 0.0366(RMS/LOC(0) - RMS/LOC) + 0.0377(\%PWR/GS(0) - \%PWR/GS)$$

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 16 may now be used in equation 14 which becomes

$$(19)\quad P = 1.4483 + 0.0351(\#TP) + 0.1765(RMS/GS) - 0.0366(RMS/LOC) + 0.0377(\%PWR/GS).$$

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:

$$(20)\quad P = 6.3750 + 0.1545(TP) + 0.7769(RMS/GS) - 0.1611(RMS/LOC) + 0.1659(\%PWR/GS)$$

A plot of this fuction versus the task difficulty, obtained from equation 10, is provided in Figure 13. 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

ORIGINAL PAGE IS
OF POOR QUALITY

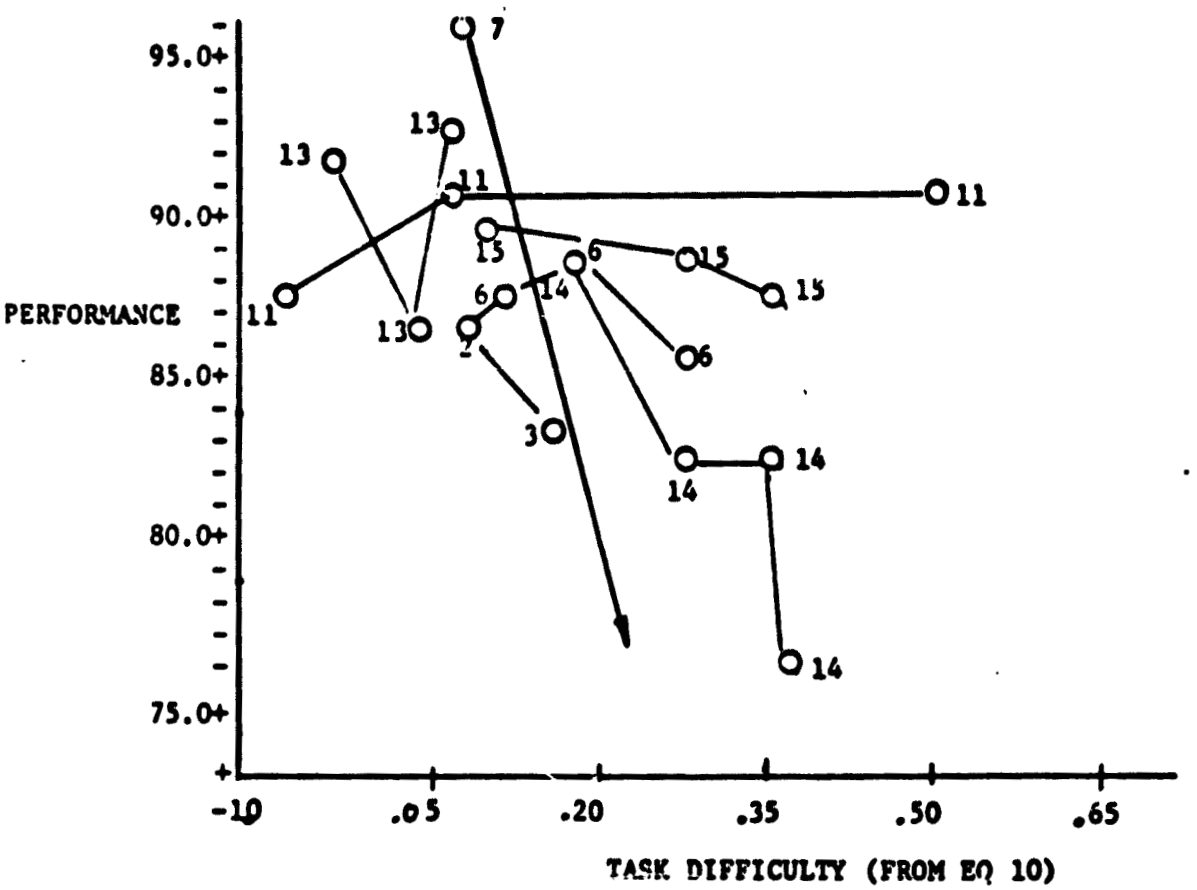

Figure 13. Combined Performar From Model) vs perceived task Task Difficulty for 7 Pilots Used in Model Development

ORIGINAL PAGE IS
OF POOR QUALITY

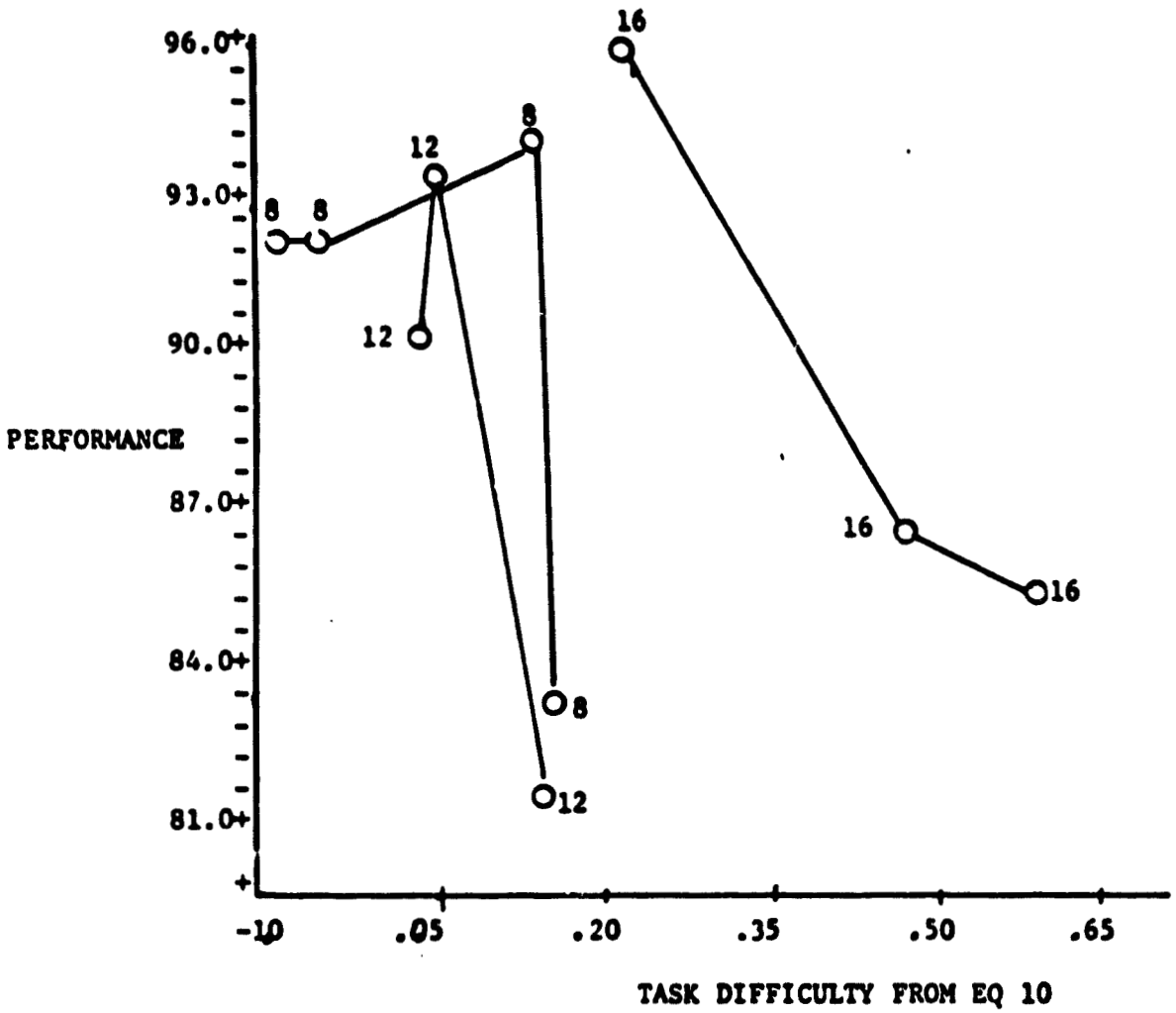

Figure14.Combined Performance vs Task Difficulty for3Test Cases of Model

ORIGINAL PAGE IS
OF POOR QUALITY

rapidly than the curves for the 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 14. 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. 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.

CONCLUDING REMARKS

Our results suggest that in a skilled task such as piloting where instrument scan plays an important role, the scanning behavior may serve as an indicator of both workload and skill. The results presented do not, at this time, seem to support the notion of an accurate, absolute measure of workload. However, a quantitative, relative comparison of mental workload under varied conditions does appear to be feasible.

One implication of the effort applies to the estimation of workload of some new procedure which may have several possible levels. In many cases, test pilots with superior flying skills are utilized in the estimation or measurement of workload. This often leads to equivocal conclusions in comparing alternative procedures or displays. The present findings indicate that different levels of loading may be difficult to measure in skilled subjects since they appear to be less sensitive to increased difficulty (see figures 1, 9, & 11). Our results imply that pilots of moderate skill are more sensitive to the verbal loading task. Thus if one is concerned with the question of the effect of changing the level

ORIGINAL PAGE IS
OF POOR QUALITY

of difficulty of some task, then as one step in the evaluation, the use of pilots of intermediate skill at several loading levels would seem appropriate since their behavior (visual scanning and performance) will be altered more as a function of the loading task than will that of more skilled pilots.

Another possible application may be the assessment of pilot skills. The work presened here suggests that there is a relationship between the scanning behavior of the pilot and his skill level. The obvious place one might use this result is in training. One may hypothesize that, as a pilot's skills develop, his visual scanning behavior will be less and less affected by non-visual increments in workload. This hypothesis is supported by a number of our findings. It appears that as skill increases, the percentage of long dwells decreases for a particular loading level. The scan pattern used during a fixed maneuver is also unaffected by verbal loading at higher skill levels, a result supported by both the frequency of usage of different instrument fixation sequences and by correlation methods. This finding might be utilized in assessing pilots' currency, competence, and level of skill; the technique might be used to pinpoint areas which may require additional training or practice.

ORIGINAL PAGE IS
OF POOR QUALITY

Publications from this research

Ephrath, A.E., Stephens, A.T., Tole, J.R., and Young, L.R., "Instrument Scan - Is it an Indicator of the Pilot's Workload?", Human Factors Soc., Ann. Meeting, Los Angeles, October, 1980.

Harris, R.L., Sr., Tole, J.R., Stephens, A.T. and Ephrath, A.E., "Visual Scanning Behavior and Pilot Workload", Aviat Psych. Symposium, Ohio State Univ., April, 1981.

Harris, R.L.,Sr., Tole, J.R., and Stephens, A.T., "How a New Instrument Effects the Pilot's Mental Workload", Human Factors Society Meeting, Seattle, Oct., 1982.

Stephens, A.T. "Visual Scanning, Mental Workload, and Performance in Aircraft Pilots", S.M. Thesis, Dept of Aero. and Astro., M.I.T., August, 1981.

Stephens, A.T., Tole, J.R., Ephrath, A.E., and Young, L.R., "Pilot Eye Scanning Behavior as an Index of Mental Loading", 8th Annual Northeast Bioengineering Conference, M.I.T., March, 1980.

Tole, J.R., Ephrath, A.E., Stephens, A.T., and Young, L.R., "Workload and Pilot Eye Scanning Behavior", Annual Manual Control Conf., M.I.T., May, 1980.

Tole, J.R., Stephens, A.T., Harris, R.L.,Sr., and Ephrath, A., "Visual Scanning Behavior and Mental Workload in Aircraft Pilots", accepted for publication in Aviat. Sp. and Environ. Med., 1981.

Tole, J.R., Stephens, A.T., Harris, R.L.,Sr., and Ephrath, A. "Quantification of Pilot Workload via Instrument Scan", AIAA Workshop on Inflight Testing of Performance and Workload, Edwards AFB, Calif., Jan., 1982.

Tole, J.R., Stephens, A.T., Vivaudou, M.B., Harris, R.L.,Sr., and Ephrath, A.E., "Entropy, Instrument Scan, and Pilot Workload", IEEE Systems, Man, and Cybernetics Meeting, Seatle, Wash., Oct., 1982.

ORIGINAL PAGE IS
OF POOR QUALITY

References

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

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

Bahill, A.T. "How Does the Saccadic Eye Movement Controller Adapt for Pathological States?", IEEE Conf on Decision and Control, New Orleans, Dec., 1977.

Bergeron, H.P. "Single Pilot IFR Autopilot Complexity/Benefit Tradeoff Study", AIAA Aircraft Systems and Technology Meeting, Anaheim, California, August, 1980, AIAA Paper #80-1869.

Burns, S.K., Dowling,N.B., Peterson,S.K., 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.

Chiles, W.D., Jennings, A.E., and Alluisi, E.A., "The Measurement and Scaling of Workload in Complex Performance", FAA Report FAA-AM-78-34, Sept., 1978.

Chiles, W.D. and Jennings, A.E., "Time-sharing ability in Complex Performance: An expanded replication", FAA Report # FAA-AM-78-33, Sept., 1978.

Clement, W.F., Graham, D., and Best, J.J., A Reexamination of Eye Movement Data, Systems Technology Inc, Tech Memo 163-A , Nov 1966, Rev. Feb, 1967.

Connelly, M.E., "Simulation Studies of Airborne Traffic Situation Display Applications - Final Report", Electronic Systems Laboratory, M.I.T., Report ESL-R-751, May, 1977.

DeMaio, J., Parkinson, S., Leshowitz, 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.

Fisher, D.F., Monty, R.A., and Senders, J.W., eds. Eye

ORIGINAL PAGE IS
OF POOR QUALITY

Movements: Cognition and Visual Perception, Lawrence Erlbaum and Associates, Hillsdale, N.J., 1981.

Gafni, H. and Zeevi, Y.Y., "A Model for Separation of Spatial and Temporal Information in the Visual System", Biol. Cybernetics, 28: 73-82, 1977.

Gerathewohl, S.J., "Optimization of Crew Effectiveness in Future Cockpit Design: Biomedical Implications", Aviat. Space Environ. Med. 47(11)1182-1187, 1976.

Gerathewohl, S.J., Brown, E.L., Burke, J.E., Kimball, K.A., Lowe, W.F., and Stackhouse, S.P., "Inflight Measurement of Pilot Workload : A Panel Discussion", Aviat. Space Environ. Med. 49(6):810-822,1978.

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.

Jex, H.R., and Clement, W.F., "Defining and Measuring Perceptual-Motor Workload in Manual Control Tasks", Systems Tech. Inc. Tech. Report 1104-1, March, 1978.

Jones, D.H., Coates,G.D., and Kirby, R.H., "The Effectiveness of Incorporating A Real-Time Oculometer System in a Commercial Flight Training Program", Final Report under NASA Grant NCCI-57, Dept of Psychology, Old Dominion Univ to NASA/Langley Research Center, October, 1982.

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.

ORIGINAL PAGE IS
OF POOR QUALITY

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.

Miller, G.A., "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", Psychological Review 63(2):81-97, March, 1956.

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

North, R.A., Harris, S.D., and Owens, J.M., "Test-retest Reliablility of Individual Differences in Dual-Task Performance", Naval Aerosp. Med. Res. Lab. Report NAMRL-1248, July, 1978.

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.

Senders, J. "Visual Scanning Behavior" in Visual Search, Proceedings of National Research Council Committee on Vision Symposium, 1970, National Academy of Science, Wash., D.C.

Senders, J.W., Elkind, J.I., Grignetti, M.C., and Smallwood, R., "An Investigation of the Visual Sampling Behaviour of Human Observers", NASA CR-434, April, 1966.

Shannon, C.E. and Weaver, W. The mathematical theory of communication, Univ. of Illinois Press, 1949.

Spady, A.A., "Airline Pilot Scanning Behavior during Approaches and Landing in a Boeing 737 Simulator", Presented at AGARD 25th Guidance and Control Panel Meeting/Symposium on "Guidance and Control Design Considerations for Low Altitude and Terminal Area Flight", Dayton, Ohio October 17-20, 1977.

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.

Tole, J.R., Stephens, A.T., Harris, R.L.,Sr., and Ephrath, A., "Quantification of Pilot Workload via instrument Scan", AIAA

ORIGINAL PAGE IS
OF POOR QUALITY

Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, Edwards, California,Jan., 1982.B

Waller, M.C., "An Investigation of Correlation between Pilot Scanning Behavior and Workload using Stepwise Regression Analysis" NASA TM X-3344, March, 1976.

Waller, M.C. and Flowers, G.S., "Dwell Time Scanning Characteristics of Pilots during a Simulated Instrument Approach", Presented at the Eye Movements and Psychological Processes, II Symposium, Monterey, California, Feb. 7-9, 1977.

Weir, D.H. and Klein, R.H., "The Measurement and Analysis of Pilot Scanning and Control Behavior during Simulated Instrument Approaches", NASA CR-1535, June, 1970.

Wittenborn, J.R., "Factorial Equations for Tests of Attention", Psychometrika, 8(1):19-35, March, 1943.