

## Quantification of Pilot Workload via Instrument Scan

J.R. Tolo(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).

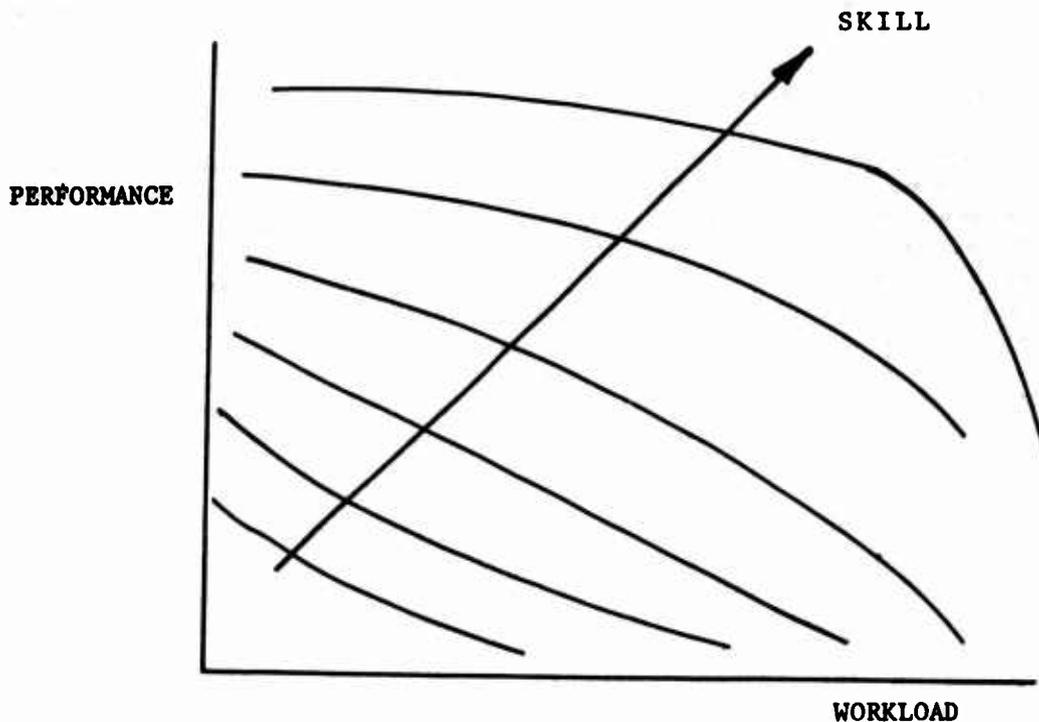


Figure 1. INTUITIVE RELATIONSHIPS BETWEEN PERFORMANCE, SKILL, & WORKLOAD

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 instrument 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 in 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 (Tolle, 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). 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. 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 maneuvers. 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:

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

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 occurrence 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 (Tola, 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 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 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.

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

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 answered 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 frequencies 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 with 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:

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

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

#### 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.

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

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  
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#	SKILL SCORE
3	100%
4	85
11	77
13	53
15	38
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 (Tolle, 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 scans were analyzed, and the percentage occurrence of the ten most frequently occurring sequences were also analyzed. These results

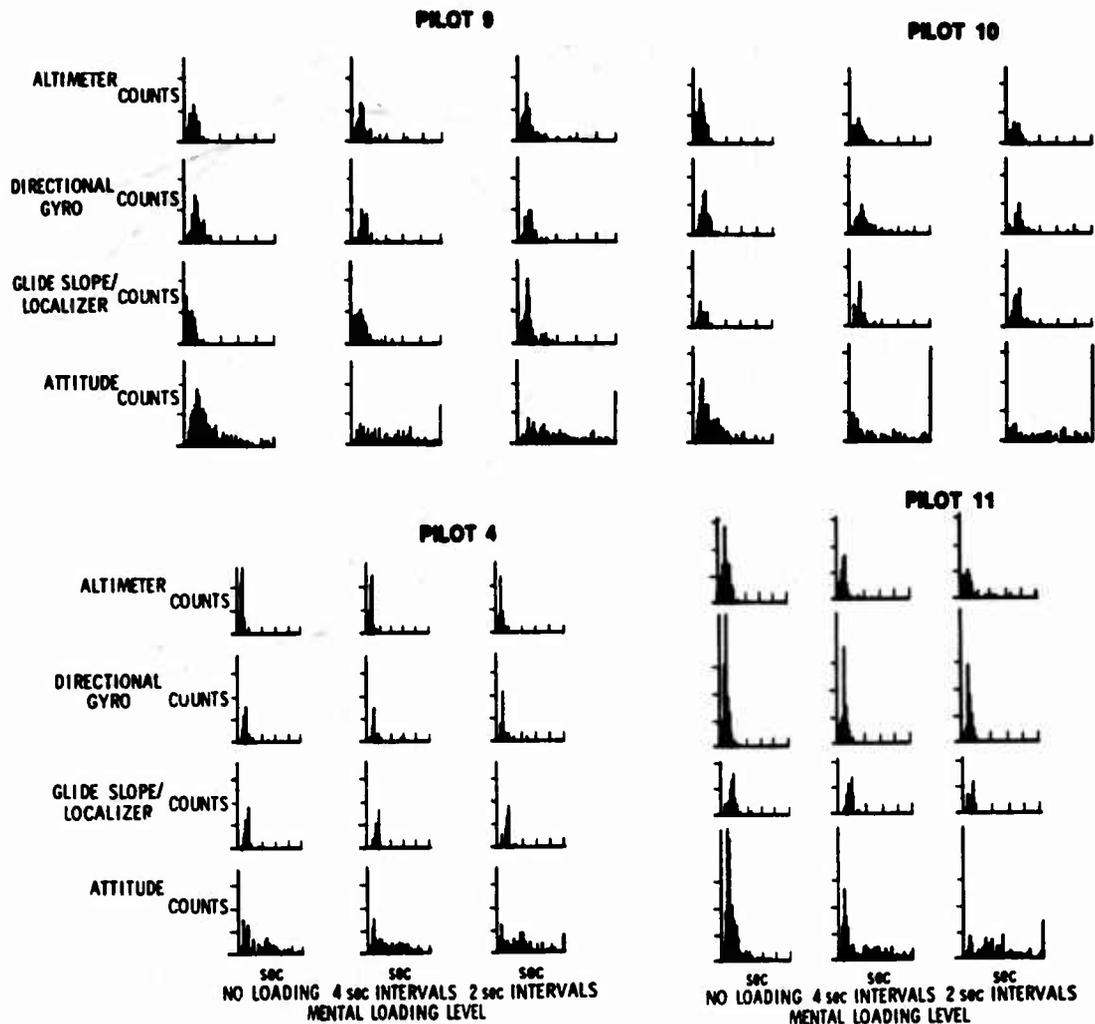


Figure 2. DWELL TIME HISTOGRAMS FOR TWO SKILLED PILOTS (#4 & #11) AND TWO NOVICE PILOTS (#9 & #10) UNDER VARIOUS LOADING CONDITIONS

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, 1978). 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 arrangement 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.

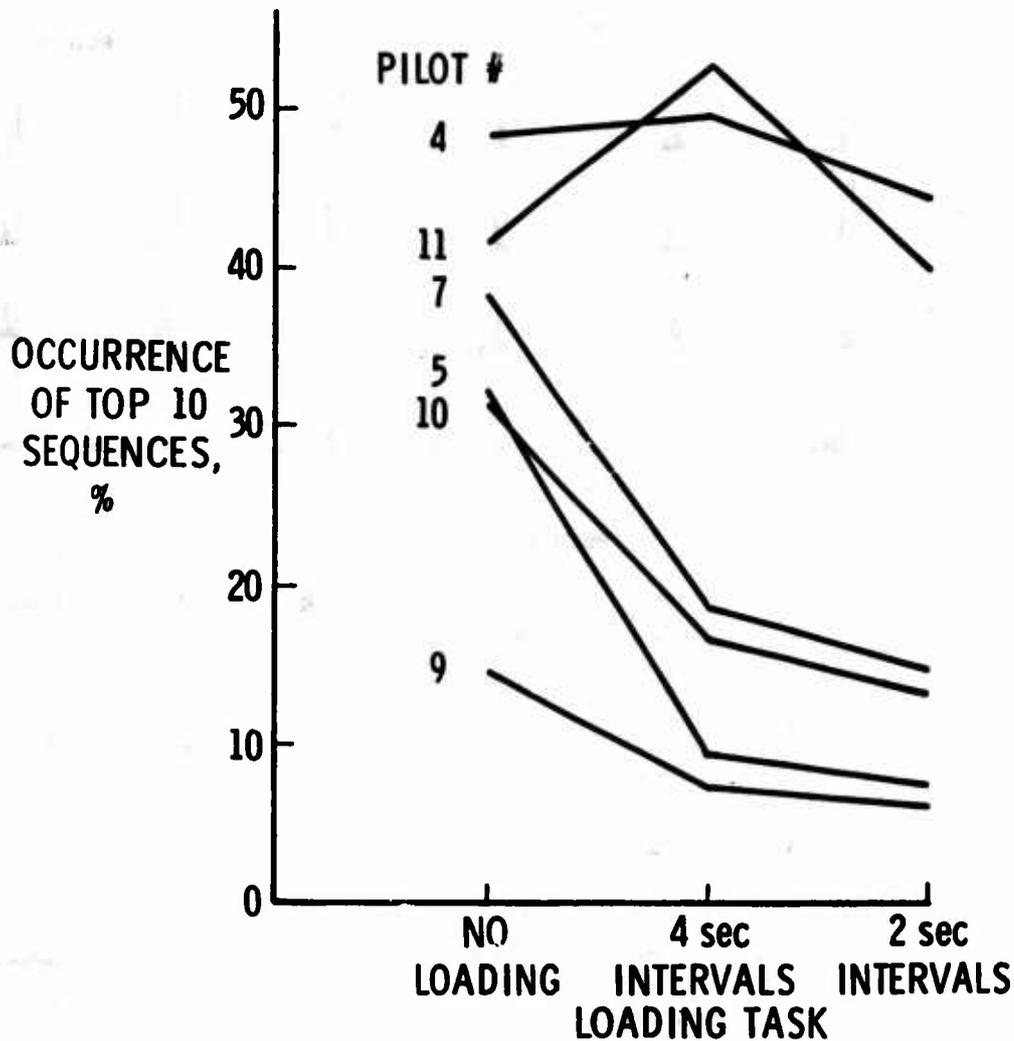


Figure 3. PERCENT USAGE OF LENGTH 4 SEQUENCES UNDER VARYING LOAD (TYPICAL SEQ : ATT - DG - ATT - ALT)

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 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):

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

where

$$H = \text{observed average entropy}$$

$$p_i = \text{probability of sequence } i \text{ occurring}$$

$$D = \# \text{ of Different sequences in the scan}$$

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 the methodology 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:

$$(6) \quad Q = \frac{M!}{(M-N)!} = \text{maximum number of sequences of length } N$$

The number of bits required to uniquely encode all Q possible sequences is  $\log_2 Q$ . The magnitude of this latter number also represents Hmax of the visual scan for the number of instruments and 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:

$$(7) H_{corr} = H_0 * \frac{H_{max}}{\log L^2}$$

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:

$$(8) H_{rate} = H_0/t$$

where  $H_0$  is the entropy for the system given by 7 and  $t$  = smallest interval in which a transition may occur.

In practice, the calculation of  $H_{rate}$  was an average value given by the following:

$$(9) H_{rate} = \frac{\sum_{i=1}^D H_{corr_i}}{DT_i}$$

where

$H_{corr_i}$  = Normalized entropy for  $i$ th sequence

$DT_i$  = Average Dwell time for  $i$ th sequence

$D$  = # of different fixation sequences

It is helpful to estimate the maximum value which  $H_{rate}$  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 Christliff, 1980) a fixation rate this high (10 fixations/sec) rapidly leads to oculomotor fatigue. A more realistic 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

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

max

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.

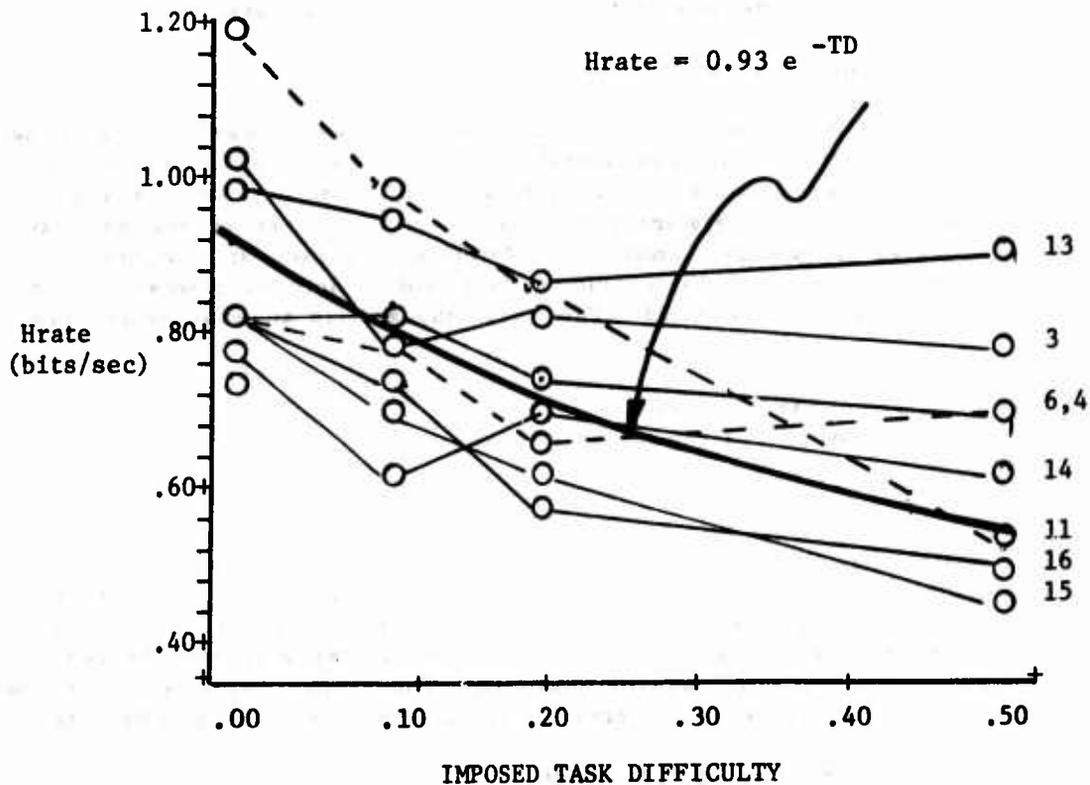


Figure 4. ENTROPY RATE ON LENGTH 2 SEQUENCES vs. IMPOSED TASK DIFFICULTY

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.

$$(10) Hrate = 0.9278 \text{ EXP}(-TD)$$

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 with the following results:

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

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 determining whether such empirical expressions are valid. The model chosen was an exponential form:

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

This equation may be rearranged as follows:

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

which states that the exponential term is equal to the difference in 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

$$(14) P(0) - P = a(TP(0) - TP) + b(RMS/GS(0) - RMS/GS) \\ + c(RMS/LOC(0) - RMS/LOC) + d(\%PWR/GS(0) - \%PWR/GS) \\ + e(\%PWR/LOC(0) - \%PWR/LOC)$$

A multiple regression analysis was then performed 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 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 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:

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

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 15 may now be used in equation 3 which becomes

$$(16) P = 1.4483 + 0.0351(\text{TP}) + 0.1765(\text{RMS}/\text{GS}) \\ - 0.0366(\text{RMS}/\text{LOC}) + 0.0377(\% \text{PWR}/\text{GS}).$$

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:

$$(17) P = 6.3750 + 0.1545(\text{TP}) + 0.7769(\text{RMS}/\text{GS}) - 0.1611(\text{RMS}/\text{LOC}) \\ + 0.1659(\% \text{PWR}/\text{GS})$$

A plot of this function 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 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 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 raising 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.

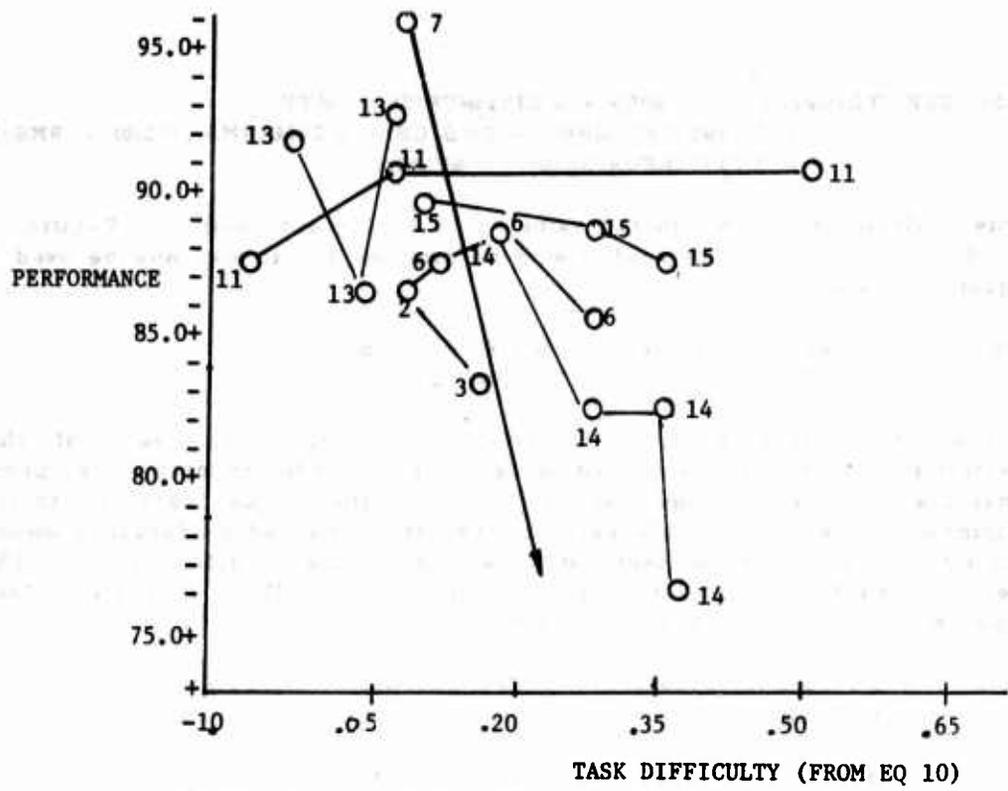


Figure 5. Combined performance (from model) perceived task difficulty for 7 pilots used in model development

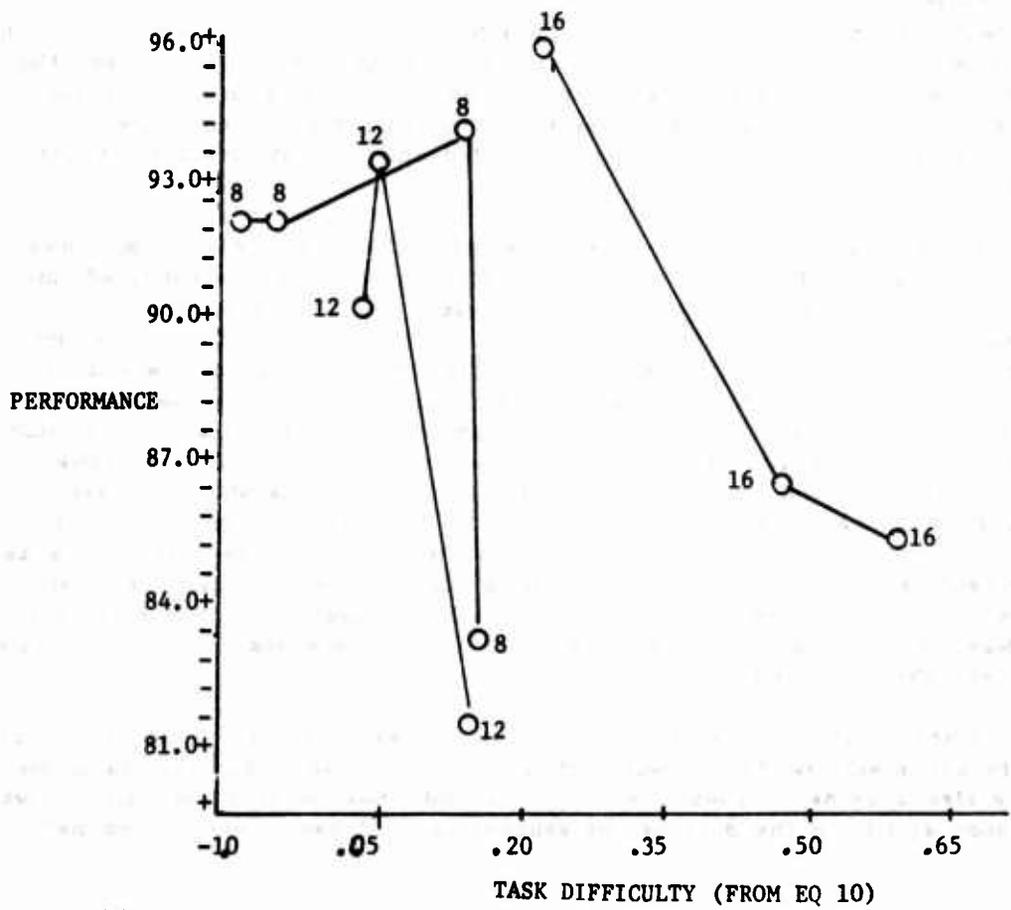


Figure 6. Combined performance vs. task difficulty for 3 test cases of model

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 from 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.

**Acknowledgements:** 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 Psychology. 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., Leshetz, 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 Nixon, 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, 118-127.

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 65886 ), 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-144851, 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-8410, 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 Control, 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 of varying difficulty) does not in itself require an altered scan path for its successful performance. As the critical task difficulty increases the dwell 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.

