General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

(NASA-TM-85145) ENTROPY, INSTRUMENT SCAN AND PILOT WORKLOAD (NASA) 8 p HC A02/MF A01 CSCL 05H

N83-19476

Unclas

G3/54 08777

IEEE Conference on Systems, Man & Cybernetics
Seattle, WA
October '1982

Entropy, Instrument Scan, and Pilot Workload

J.R. Tole[1,3], A.T. Stephens[1,4], M.Vivaudou[3], R.L. Harris,Sr.[2], and A. Ephrath[5]

[1]Harvard/MIT Division of Health Sciences and Technology [2]NASA Langley Research Center
[3]Current Address: Biomedical Engineering Dept., Worcester Polytechnic Institute, Mass.
[4]Current Address: Boeing Co., Renton, Wash. [5]Bell Labs, Piscataway and M.I.T.

MAR 1983
RECEIVED
NASA STI FACILITY
ACCESS DEPT.

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 paper describes work in progress on the use of techniques based on correlation and information theory to analyze the relationships between mental loading and visual scanpath of aircraft pilots. It is part of a more general investigation of the relationship between skill, performance, mental workload, and visual scanning behavior.

The experimental method 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.

Analysis of the periodicity of the subject's intrument scan was accomplished by means of correlation techniques. For skilled pilots, the autocorrelation of instrument/dwell times sequences showed the same periodicity as the verbal task. A similar result was not observed in unskilled pilots. This finding suggests that ability to multiplex simultaneous tasks increases with skill. Thus autocorrelation may provide a way of evaluating the operator's skill level.

The concept of entropy from information theory was employed to estimate changes in scanning pattern as a function of verbal task difficulty. A normalized expression for the entropy rate of a sequence of instrument fixations was designed and its computation for small-length sequences showed a consistent decrease with increasing loading levels. Furthermore task difficulty and entropy rate were found to be exponentially related, therefore suggesting the utility of entropy rate of instrument scan as a tool for mental workload assessment.

Introduction

The quantification of mental workload in aircraft pilots has been of considerable interest for some ime. 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 because of its interdependence with variables such as pilot skill. A previous study (Tole et al, 1982.B) suggested that the intuitive relationship between skill, performance, and mental workload illustrated on figure 1 can be quantitatively tested by combining a measure of workload based on instrument scan with independent measures of pilot skill and performance. Based on the same hypothesis, this research demonstrates the use of simple mathematical tools to extract quantitative information from scanning patterns.

ORIGINAL PAGE IS
OF POOR QUALITY

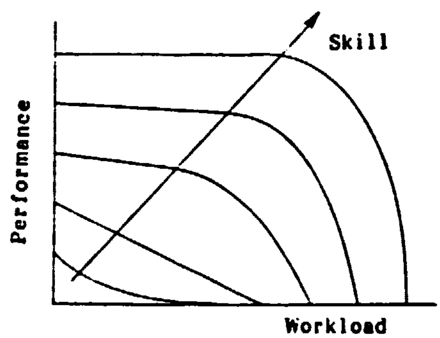


Fig.1.Theoretical relationship between performance, skill, & workload.

Visual Scanning Behavior

The pilot has many sources of information input but the most important one during instrument flight is probably the visual pathway. 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.

ORIGINAL PAGE IS
OF POOR QUALITY

Experimental Design

A series of experiments is being carried in order to carefully examine these ideas. The basic experiment is described in detail elsewhere (Tole,et al,1982.A) 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:

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

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. Calibration using a side task (Ephrath, 1975) confirmed the relative difficulty of these number intervals.

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

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.

Results

Instrument dwell time histograms and the frequency of usage of different sequences of instrument fixations were both affected by the loading task. Both results are reported in detail elsewhere (Tole, et.al., 1982.B) and only the major points are mentioned here. Piloting and number task performance were evaluated and recorded and a combined performance measure was computed. Skill was estimated independently via a method based on pilot experience (Hollister, et al, 1973). The results indicated an increase in fixation dwell times, especially on the primary instruments with increased mental loading. Skilled subjects "stared" less under increased loading than did novice pilots. Also the percentage occurrence of the subject's most used sequences decreased with increased task difficulty for novice subjects but not for highly skilled subjects.

Quantifying the scan

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

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 communication 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):

$$H_o = -\sum_{i=1}^{D}[p_i \log_2 p_i] \qquad (2)$$

ORIGINAL PAGE IS
OF POOR QUALITY

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.

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).

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) 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, 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:

(3) $$Q = M \cdot (M-1)^{N-1} = \text{maximum number of sequences of length } N$$

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 7 instruments the value of Q for sequences of 2 instruments is 56 which yields a corresponding $H_{max} = 5.8$.

The normalized value of H may then be calculated from:

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

where
$A = \log_2 L$ 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)

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:

(5) $$H_{rate} = H_o/t$$

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

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

ORIGINAL PAGE IS OF POOR QUALITY

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

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 statistics for the various instrument sequences in the scan. While it is possible for pilots to make rather rapid glances (with dwell times of 100 msec or less) at their instruments (Harris and Christhilf, 1980) a fixation rate this high (10 fixations/sec) rapidly leads to oculomotor fatigue. A more realistic average value is probably about 2 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 2 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 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 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 2 plots H_{rate} vs number task difficulty for several pilots.

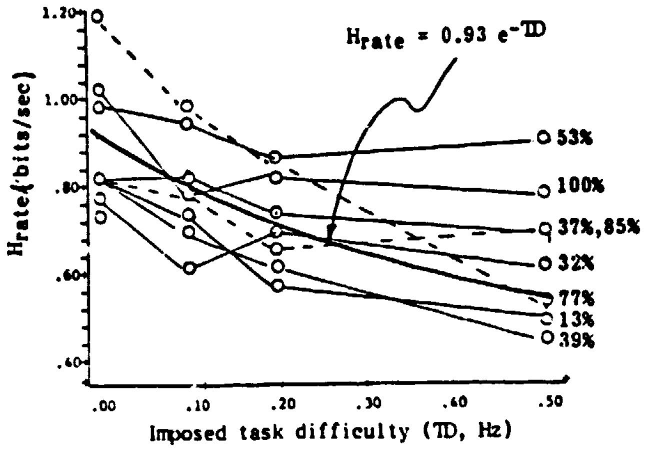


Fig.2.Entropy rate on length-2 sequences vs imposed task difficulty for 8 pilots (relative skill levels shown on the right - highest=100%).

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 9 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 6 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).

ORIGINAL PAGE IS OF POOR QUALITY

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

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

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:

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

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. A sequence of instrument numbers versus time was developed from the data and stored on a disk. 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:

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

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

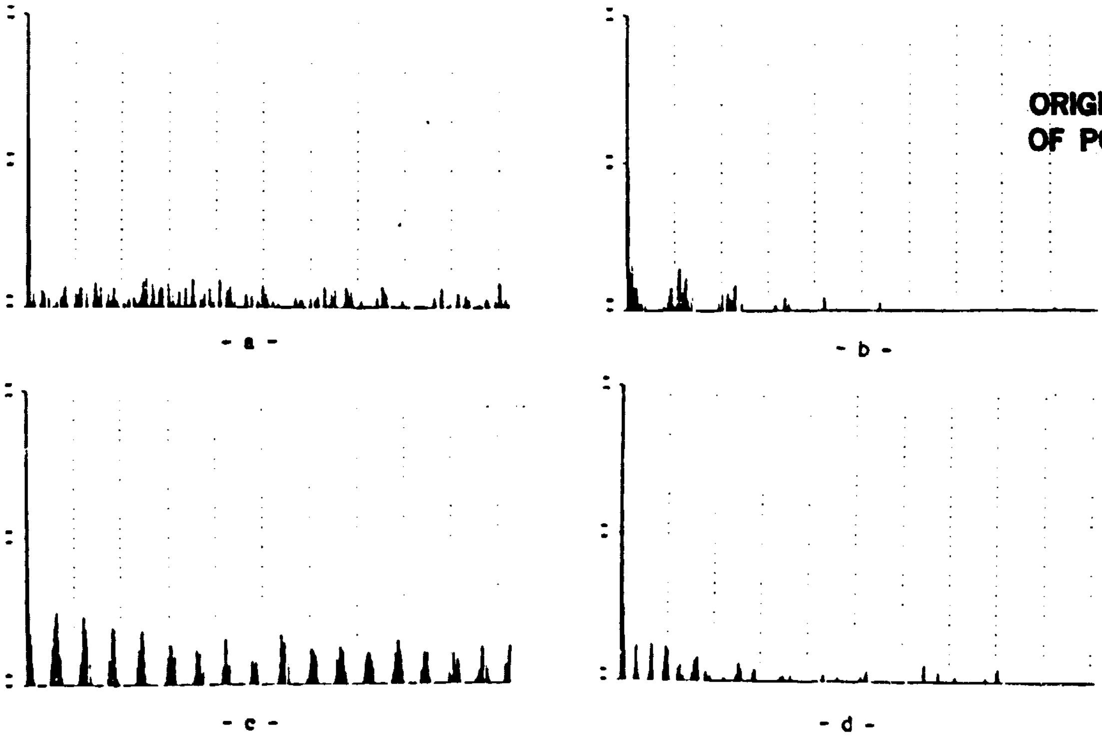
The sample autocorrelation of $\{f_j(i)\}$, or sample autocovariance of $\{x_j(i)\}$, was calculated by the formula:

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

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 Figure 3. 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 Figure 4 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.4.a). When the periodic task is applied, the scan becomes more and more periodic with increased task frequency (cf fig.4.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



ORIGINAL PAGE IS
OF POOR QUALITY

Fig.3 Autocorrelations for pilot #4 (relative skill level = 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

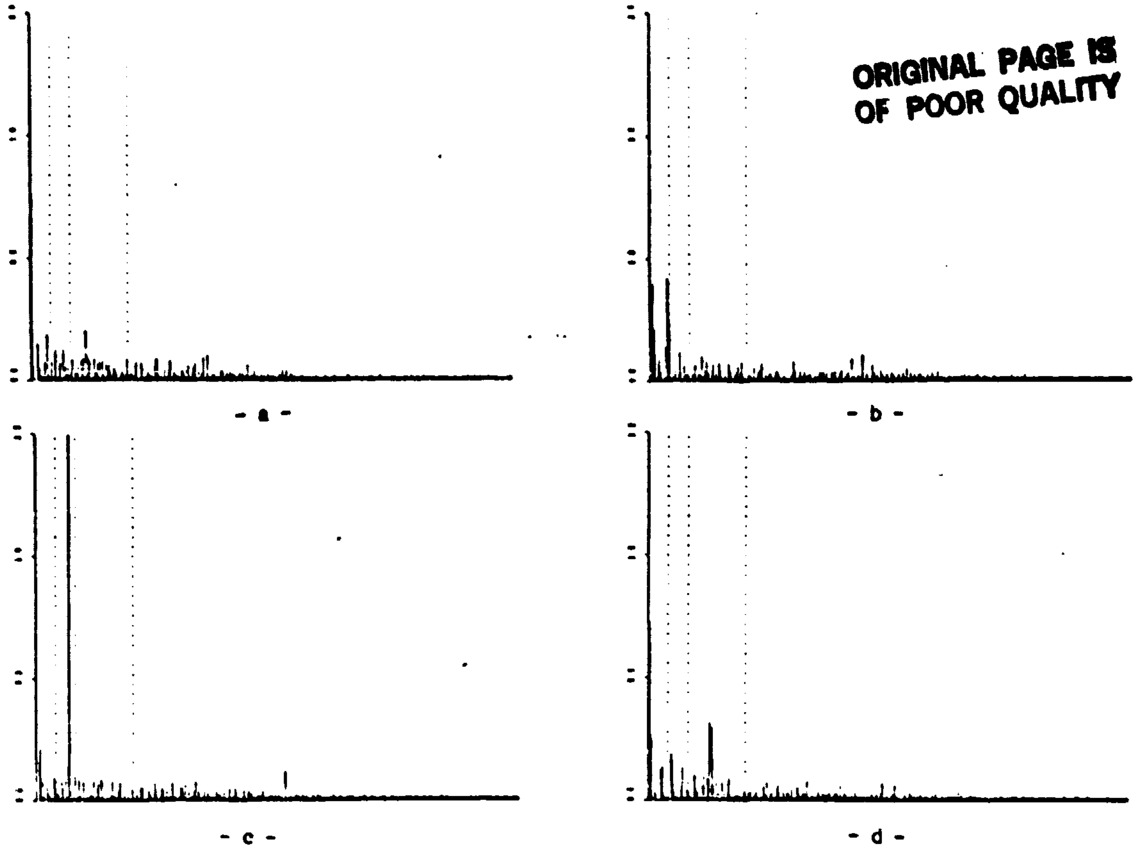


Fig.4.Power spectral densities for pilot #4 (relative skill level = 85%) using attitude indicator (dotted lines correspond to frequencies of 0.1, 0.2, and 0.5 Hz respectively). Number task intervals and associated task difficulties are a) no intervals - 0, b) 10 sec - 0.1, c) 5 sec - 0.2, d) 2 sec - 0.5.

(cf fig.3.d). A similar behavior is observed for all of the higher skilled pilots as demonstrated in Table I. The periods of oscillation for the 5 pilots of highest skill appear to match those presented to them by the number task very closely. However, the other 6 pilots do not seem to have any consistent pattern in their autocorrelation of sequences. Most of the 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 (1976) 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.

Pilot's relative skill levels	Number task periods(sec) 11	6	3
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 1. Scan autocorrelation dominent periods for 9 pilots using attitude indicator (glide slope/ localizer for *) for 3 frequencies of the mental loading task.

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 indicated that pilot's scanning patterns are significantly affected by the level of mental workload in a manner dependent on skill.

In order to quantify these alterations of the scan, two methods were investigated and are described here: a) Entropy - Under relatively constant instrument flight conditions, the entropy rate of the visual scan path has been found to be exponentially related to the