FMP STUDY OF PILOT WORKLOAD
Quantification of Workload via Instrument Scan

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

[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]Bell Labs, Piscataway and M.I.T.

Introduction - Visual Scanning Behavior

The aircraft 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 method discussed here 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, fatigue, etc. 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.

One of the main reasons for measuring mental workload is to predict situations under which piloting performance will fall below an acceptable (safe) level. A persistent difficulty in workload measurement appears to be the influence of skill on the performance vs. workload relationship as noted graphically in figure 1. Skilled pilots tend to maintain a high level of performance under increasing mental load until a precipitous performance decrement occurs during an overload situation. Our work to date has supported this notion and led us to explore the relationships suggested in figure 1. Space does not allow a detailed discussion of this issue here; it is sufficient to note that the measures we are using are dependent on subject skill (indeed the measures may eventually provide good indicators of skill) and that novice pilots may be more desirable subjects if the effect of workload on performance is to be explored over wide ranges of task difficulty.

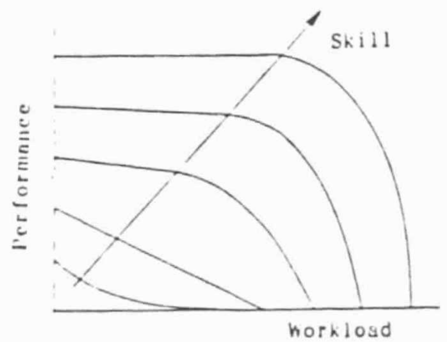

Fig.1.Theoretical relationship between performance, skill, & workload.

Methodology

At present, our technique requires a piloting scenario which forces a relatively stereotyped instrument scan pattern and an independent verbal (or possibly visual; see below) task of varying difficulty. The following is an abbreviated description of our previous experiments which should serve as a general orientation to the methodology. Details of the methods, including data analysis techniques are discussed elsewhere (Tole, et al, 1982a & b).

Pilot subjects are asked 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 increases with frequency is used to increment the subject's mental workload. 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 (Middleton, et.al., 1977) mounted so as not to interfere with the subject's view of the instrument panel. This device is non-invasive and provides the time course of the pilot's instrument fixations 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 requires the pilots to respond to a series of evenly spaced three-number sequences (Wittenborn, 1943) presented audibly. The pilot was told that he must respond to each 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). 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 mental workload experienced by the pilot is inversely proportional to the intervals between number sequences. This relationship is given by the following arbitrary equation :

(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 hare 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.

Several variables are obtained from each of the two tasks in order to compute 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.

Performance on the piloting task was estimated from the glide slope and localizer errors. In addition to the usual measure of total RMS error from the intended flight path, a measure of "smoothness" of ride, estimated from the frequency content of oscillations about the intended path was also included. 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. Combining the % of the spectral power above 0.1 Hz for the glide slope and localizer with the RMS errors for these instruments yields a performance score for the piloting task. This measure may be combined with the score from the number task to obtain a total performance score.

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 (e.g. total flight time, total time in type, yrs. since certification, etc.) was used for this purpose (Hollister, et al, 1973).

Effects of Loading on the Visual Scan

Instrument dwell time histograms and the frequency of usage of different sequences of instrument fixations were both affected by the loading task. Pilots tend to stare at the primary instrument (Attitude indicator in our experiments) as the level of difficulty of the verbal task increases. The percentage usage of various fixation sequences (eg. ATT-DG) also decreased with increasing task difficulty. Both of these trends were much more prevalent in novice subjects as compared with skilled pilots.

The piloting task in the basic 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):

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

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

(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 $\log_2 Q$. The magnitude of this latter number also 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 $H_{max} = 5.8$. In order to include the effect of instrument dwell times in our measure, a term for entropy rate was defined as:

(4) $$H_{rate} = \sum_{i=1}^{D}[H_i/DT_i]$$

where H_i = entropy for ith sequence
DT_i = Average dwell time for ith sequence
D = Number of different fixation sequences

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:

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

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 basic 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 our test subjects.

This data is modeled by expressing H_{rate} as an exponential function of TD.

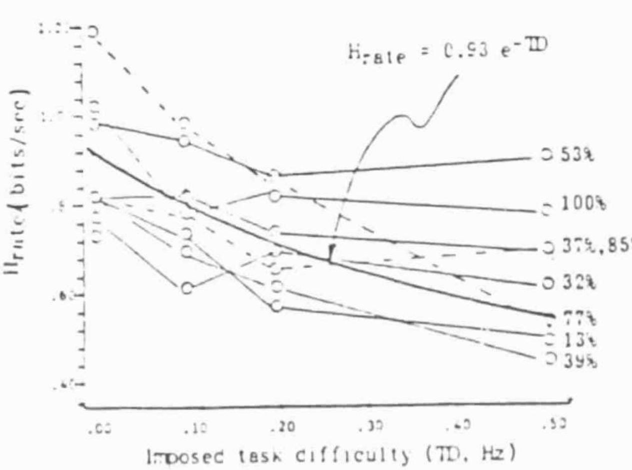

Fig.2.Entropy rate on length-2 sequences vs imposed task difficulty for 8 pilots (relative skill levels shown on the right - highest=100%).

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

This equation may be solved for task difficulty to yield

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

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

Autocorrelation

The relationship between instrument scan frequency and number task presentation frequency provides valuable insight on how the task, and therefore the associated mental load, affects the scanning pattern. Due to the periodic nature of the verbal task, the use of autocorrelation to analyze the scanning behavior is one possible method for examining this relationship.

Autocorrelation was performed on the scanning data as follows. 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 $\{x(i)\}$ where the value is 1 for the instrument being studied and 0 for all other instruments. In order to eliminate the dc component for later spectral analysis, a zero-mean sequence $\{f(i)\}$ was computed from $\{x(i)\}$ 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. An example of this analysis is

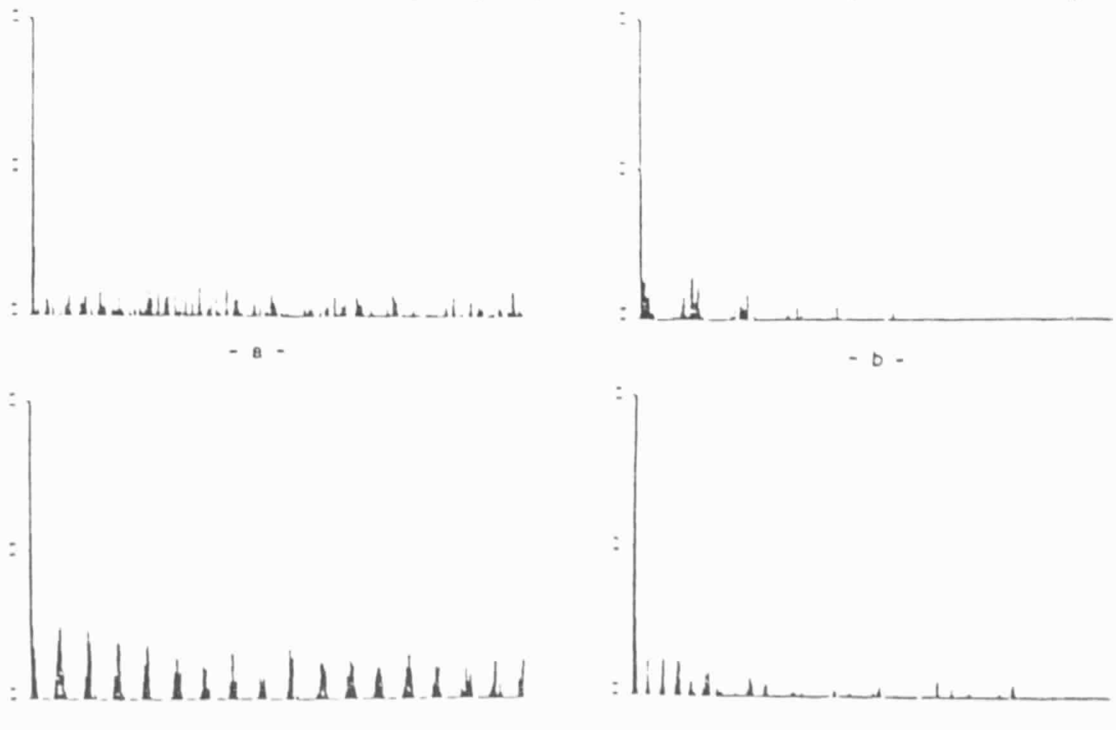

Fig.3 Autocorrelations for pilot #4 (relative skill level = 85%) using attitude indicator (dotted lines indicate 10-sec intervals). Number task intervals and associated task difficulties are a) no intervals - 0, b) 10 sec - 0.1, c) 5 sec - 0.2, d) 2 sec - 0.5.

shown 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 density calculations 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. When the periodic task is applied, the scan becomes more and more periodic with increased task frequency. 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.

A similar behavior has been observed for all of the highly skilled pilots who have participated in our experiments thus far. Novice pilots, however, do not seem to have any consistent pattern in their autocorrelation sequences. Most of these pilots showed little or no periodicity in their scans for any of the loading conditions. One explanation may be that skilled pilots have a better developed ability to time multiplex several simultaneous tasks.

Visual Scanning Measures Applied to the FMP Flight Task I, Manually Flown ILS Approach and Landing of Two-Pilot Passenger Jet Transport

We now briefly discuss the application of our techniques to the proposed study of workload during an ILS approach. Two or three factors must be manipulated to use our techniques: 1) a piloting task requiring a stereotyped scan path, 2) a verbally presented mental loading task, or 3) a visually presented mental loading task. It is assumed that the cockpit to be used for the experiments may be outfitted with the NASA Langley oculometer system or an equivalent and that ample time will be allowed (approximately 5-10 minutes) for calibration of the oculometer before an experimental session begins.

The proposed ILS approach scenario requires the use of a stereotyped scan path, though it should be emphasized that the task and hence the scan pattern is not constant throughout the scenario. (Recall that our earlier experiments forced a repetitive scan pattern to be developed over a long constant flight maneuver) Thus, the second to second level of loading due to the flight task and the corresponding instrument scan will vary, albeit in a somewhat predictable fashion. We believe it is important to note that the difficulty of the ILS approach changes somewhat from instant to instant and almost certainly implys higher workload as the runway threshold is neared. The additional verbal or visual loading task serves to "bias" the total amount of mental load on the pilot with the goal of locating peaks in the load due to the piloting task alone. The notion here is that the workload due to the additional task is roughly additive with the instantaneous load due to the piloting task. The hope would be to bias the total load to a high enough level to demonstrate a performance decrement (which may be a non-linear function of loading) while at the same time hopefully observing a monotonic change in the measures of scanning behavior as a function of the increased load.

Several levels of difficulty of the additional task are required. These may achieved in two ways. A constant level of difficulty may be imposed over the entire approach; this method is to be recommended at present as we are not as yet sure how to analyze short segments of the scan pattern. Each level of difficulty of the imposed extra task would thus require a separate run. Since both the verbal and visual tasks are periodic, their respective difficulties may be altered during a run by changing the period between presentations of the task. This method would seem more attractive if the piloting task were indeed fixed over the entire run.

The verbal task described above may be used as one means of biasing the loading level. This has been shown to work well in our experiments and is easy to implement and score. Its limitation is that it is not a task which would ordinarily be performed in the course of flight. It would be possible to modify the task to make it more like either a constant rate of radio communication or a rapidly updated manual computation of navigational coordinates.

An alternate, visual version of this task is also possible and perhaps more appropriate for actual flight conditions. A small display could be mounted in a convenient point in the pilot's visual field. The display could present either a "+" or a "-" sign. At periodic intervals an auditory "beep" would signal that the pilot should observe this display and indicate (optionally) via a rocker switch whether the display is currently indicating + or -. The interval between "beeps" determines the difficulty of this task and one possible measure of workload is the % of time the pilot is actually able to observe the display.

Entropy rate calculations could be made on the scanning data regardless of whether the visual or verbal loading task is used. Since both tasks are periodic, the autocorrelation technique may also be applied. Although we have not done it as yet, we expect that cross correlating the time of presentation of the imposed task with the scanning data is likely to yield good results especially in the type of flight scenario proposed in this study. We expect that a characteristic "signature" will appear in the crosscorrelation between the loading task and the instrument scan and that this signature will be altered via changes in task difficulty.

Limitations and Pitfalls of the Technique

There are a number of potential problems in applying our techniques. These are ennumerated below :

1. The piloting task being performed must require instrument scan.
2. The scan must be repetitive, at present, though we are working on methods (e.g. cross correlation) for analyzing short segments of a scan pattern.
3. An onboard oculometer required and must be mounted in instrument panel (NASA - Langley Research Center has worked out many of the technical problems however).
4. It may be necessary to calibrate without the pilot's cooperation due to time limitations in the proposed experiments.
5. The behavior of the various measures of scan has not been examined under a wide variety of situations as yet, hence we are unable to comment on flight scenarios in which the task is most applicable other than the obvious requirement of some type of scanning behavior.

Despite these potential limitations, we are confident enough about the methodology that we believe it should be included as one of the techniques (hopefully in conjunction with other) in the proposed study.

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

References

1. Atteneave, F. Application of Information Theory in Pyschology, Holt, Rinehart, and Winston, 1959.
2. DeMaio, J., Parkinson, S., Leshotz, B., Crosby, J., and Thorpe, J. Visual Scanning: Comparisons between Student and Instructor Pilots, Air Force Human Factors Resources Laboratory Report AFHRL-TR-76-10.
3. Ephrath, A.E. Pilot Performance in Zero-Visibility Precision Approach, Ph.D. Thesis, Dept. of Aero. and Astro., M.I.T., June, 1975.
4. Harris, R.L., Sr., and Christhilf, D. "What do Pilots See in Displays?", presented at the Human Factors Society Meeting, Los Angeles, October, 1980.
5. 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.
6. 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.
7. 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.
8. Schwartz, M., Information Transmission, Modulation, and Noise, McGraw Hill, New York, 1959.
9. Shannon, C.E. and Weaver, W. The mathematical theory of communication, Univ. of Illinois Press, 1949.
10. 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.
11. 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.B
12. Tole, J.R., Stephens, A.T., Vivaudou, M., Harris, R.L.,Sr., and Ephrath, A., "Entropy, Instrument Scan, and Pilot Workload", IEEE Conf on Systems, Man, and Cybernetics, Seattle, Wash., Oct., 1982.
13. Wittenborn, J.R. "Factorial Equations for Tests of Attention", Psychometrika, 8(1):19-35, March, 1943.