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)
ABSTRACT

This paper describes an experimental paradigm and a set of results which demonstrate a relationship between the level of performance on a skilled man-machine control task, the skill of the operator, the level of mental difficulty induced by an additional task imposed on the basic control task, and visual scanning performance. During a constant simulated piloting task, visual scanning of instruments was found to vary as a function of the level of difficulty of a verbal mental loading task. The average dwell time of each fixation on the pilot's primary instrument increased as a function of the estimated skill level of the pilots, with novices being affected by the loading task much more than the experts. The results suggest that visual scanning of instruments in a controlled task may be an indicator of both workload and skill.
INTRODUCTION

A persistent difficulty in the design and evaluation of man-machine systems is the inability to accurately measure operator workload under various conditions. Some of this difficulty is due to an inability to define clearly what is meant by workload. So far, not a single definition of workload has been developed. This is not surprising since the topic is heavily related to the study of human behavior which itself is not an exact science. The term workload has been constructed to be descriptive of the difficulty of performing a task. One must usually rely on various descriptions of workload including verbal and graphical analogies. There are at least two components to workload: mental and physical. A list of some of the aspects of these two types is illustrated in figure 1.

The desire to measure workload is usually motivated by the need to predict situations in which operator performance will decline. The reasons for this need are evident: if the operator has too many tasks to accomplish in too short a time, the performance on all or some of the tasks may be diminished. The same may be true if the operator has too few tasks to perform and allows his attention to wane. For example, recent experiments (1) suggest that a general aviation pilot flying a simulator equipped with an autopilot has a decreased ability to detect his own blunders as the sophistication of the autopilot increases. It should also be noted that another important potential cause of performance decrement is the occurrence of an extremely rare or novel event or series of events. Pilot training methods attempt to take some of the more common of these rare events into account by having the pilot practice procedures for dealing with problems such as engine out, stall, loss of one or more instruments, etc. It may be the unusual failures which have never been seen before which represent the greatest difficulty since they may cause the pilot to focus his attention too narrowly, perhaps forgetting about his primary piloting task at a critical moment.

One of the important human factors questions in the cockpit is how to specify the procedures, displays, tasks, etc. in such a way that severe over or underloading of the pilot will not occur in any of the anticipated circumstances. Thus, if one has a choice of 5 different procedures for accomplishing some task, it would be quite useful to compare the relative difficulty of the procedures and the effects of various perturbations or external disturbances on each procedure. Were a
quantitative comparison possible, the selection of the “best” candidate procedure (or display format, etc.) might be greatly facilitated. Such measures could also be used to compare what is currently in use with an alternate approach.

Since the goal of workload measurement is the prediction of performance, it is often suggested that performance is the parameter which should be measured as the workload conditions are varied. Certain performance criteria may be set and when the pilot cannot meet them the level of loading may be judged to be too high. Such a technique assumes that performance varies in a consistent fashion with workload and skill. That is, for this approach to be generally useful, all pilots should experience about the same performance decrement for the same increase in workload. Experience suggests that this is not the case however. The point is that in situations such as piloting, where performing manual dexterity and verbal or mental activities simultaneously are especially important, performance of a skilled operator may not show a great decrement until the workload is severe, and then a precipitous decline in performance may occur.

Figure 2 is a graphical statement of our hypothesis for the relationships between workload, performance, and skill. This hypothesis is specifically directed to the high workload situation. Performance may also decrease at low workloads. This figure does show that performance remains constant over a range of workloads regardless of skill level. However, the more skilled operator can maintain that performance level at higher workloads than the less skilled. In an attempt to confirm this hypothesis,

we have been exploring these relationships by examining the behavior of aircraft pilots under varying task difficulty. In the work described here we are concerned with the variation of a dependent variable, the visual scanning of instruments, as a function of skill level, inherent task difficulty, and the difficulty of an additional verbal mental loading task. We will explore how the timing of fixations on various instruments varies as a function of verbal task difficulty and the skill level of the pilot. We will discuss the implications in the results for the assessment of skill level in a task which requires skilled visual performance under varying task difficulty and in the evaluation of learning behavior in this type of task.

DESCRIPTION OF EXPERIMENTAL PROCEDURES

These experiments are concerned with relationships between “steady-state” levels of the various independent parameters: piloting performance, skill, and workload. The approach was to attempt to demonstrate whether consistent steady state effects of a constant mental loading condition could be observed. Thus, the piloting task and verbal mental loading task were held constant for a period sufficiently long enough to collect the data to evaluate the average effects of these conditions. A run length of ten minutes was chosen as an estimate of the minimum amount of time required to provide a sufficient number of fixations to satisfy the assumption of steady state conditions. The piloting task chosen was a precision straight and level path with zero degree glide slope and live localizer with constant sensitivities on the needle movements while maintaining a constant heading and airspeed. In
order to force some pilot vigilance on this task, a low level of
turbulence was also introduced for each run.

A desktop general aviation instrument flight simulator was used to
simulate these flight maneuvers. This simulator is shown in figure 3.
Pilot lookpoint on seven instruments (Attitude Indicator, Directional
Gyro, Altimeter, Vertical Speed Indicator, Airspeed, Turn and Bank, and
Glide Slope/Localizer) was measured and recorded as indicated below.

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 in order to accomplish the
task) while providing constant mental loading during the maneuver. This
was accomplished by having the pilots respond verbally to a series of
evenly spaced three-number sequences (4) presented to them by a tape
recorder. The pilot was told that he must respond to each three-number
sequence as either "plus" or "minus" (up or down respectively on a rocker
switch) according to the algorithm: first number largest, second number
smallest = "plus" (e.g., 5-2-4), first number smallest, third number
largest = "plus" (e.g., 1-5-7), all other sequence combinations are
"minus" (e.g., 9-5-1). The numbers were prerecorded at 4 second and 2
second intervals between sequences. These spacing intervals were
determined empirically to create heavy and severe additional mental
loading respectively. The pilot was instructed to give the number task a
priority equal to that of the piloting task (as if the verbal task
represented a constant rate of radio communication).

Pilot lookpoint was measured using a Honeywell oculometer system
which has been substantially modified by NASA Langley Research Center
(3). 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 second. Starting with
this information, dwell time histograms for each instrument for each
loading condition could be plotted as discussed below. These histograms
are a plot of the number of fixation counts which fall into bins of
specified time duration during a run. In these figures, the bin size is
0.066 seconds and the range of the time axis is 0 to 5 seconds. The
Instruments name for each histogram is located on the left side of the
figure.

Six subjects, varying in skill level from non-pilot to a highly
experienced NASA test pilot participated in these experiments. The subject numbers and approximate skill level are listed below.

<table>
<thead>
<tr>
<th>Subject Number</th>
<th>Skill Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Highly skilled NASA test pilot</td>
</tr>
<tr>
<td>11</td>
<td>Highly skilled general aviation Instructor pilot, NASA employee experienced in simulation experiments</td>
</tr>
<tr>
<td>9</td>
<td>General aviation pilot, current only in simulators</td>
</tr>
<tr>
<td>5</td>
<td>General aviation pilot, current in airplanes</td>
</tr>
<tr>
<td>10</td>
<td>Student pilot</td>
</tr>
<tr>
<td>7</td>
<td>Non-pilot</td>
</tr>
</tbody>
</table>

Subjects 5, 7, 9, and 10 are referred to collectively as "novice" pilots in this paper. Subjects 9 and 11 were the only pilots with any previous experience on this particular simulator. The subjects were allowed to practice the flying maneuver and verbal mental task until they felt comfortable with the situation.

RESULTS

Dwell Time Histograms

Figure 4.— Dwell histograms for pilot 5

Figure 5.— Dwell histograms for pilot 11

Perhaps the most striking effect observed in these experiments is the effect of the verbal loading task on the dwell time histograms of individual instruments for a given maneuver. In the four novice subjects, the dwell time on the primary instrument (the attitude indicator in all but the non-pilot who used the glide slope/localizer) became progressively weighted toward extremely long dwells as the verbal task difficulty increased. Figure 4 shows the dwell time histograms for pilot 5 on the attitude indicator, directional gyro, glide slope/localizer, and vertical speed indicator. Note that for the no mental loading case, the dwell histogram on the attitude indicator has a fairly standard shape (2). When numbers are added to the piloting task, the dwell becomes longer and the mode of the histogram at 1/2 second begins to disappear. The effect is even more dramatic for the 2-second interval case; the entire distribution is skewed toward extremely long dwells on attitude as the pilot apparently begins to "stare" more and
more at this instrument. Similar effects are seen for the other novice pilots.

However, an interesting difference occurs for subject 7, the non-pilot. This subject had no previous piloting experience and was only given enough practice to allow him to stay nominally on course during the precision straight and level maneuver. This subject adopted the glide slope/localizer as the primary instrument apparently in an effort to accomplish the precision task by keeping the needles centered. Even though the subject adopts the inappropriate instrument to accomplish the piloting task, the dwells on this instrument are affected in a manner similar to those on the attitude instrument for the more experienced novice subjects.

The visual scanning behavior of the two subjects with higher levels of skill was also affected by the verbal loading although to a much lesser degree than for the novice pilots. Figure 5 shows the dwell time histogram for subject 11, who had the next to highest skill level, and was somewhat more affected than the test pilot, especially at the highest loading level. Subject 11 uses a large number of short dwells on the attitude indicator under the no loading case. When the mental loading task is introduced at 4-second intervals, his distribution is shifted to somewhat longer dwells. However, there is still a very significant peak at around 1/2 second. The actual shift in dwell times is not as large as that seen in the novice pilot's histograms, even though there appears to be a large change due to the reduction in magnitude of the histogram peak. Even at the highest mental loading level, the shift to longer dwells is not as severe as it is in the less skilled pilots.

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

The dwell time characteristics on secondary instruments were affected most in the novice subjects, as may be seen in figure 4. This is typical of all novice subjects. The secondary instrument dwells are seen to change in a different manner than the primary instrument dwells. As opposed to the shift to longer dwells, as in the case for the primary instruments, the effect of mental loading is to decrease the number of looks at secondary instruments, one example of a phenomenon known as load shedding. The shape of some of the histograms changes under varying loading conditions. Pilot 4 was the only subject whose dwell time histograms on secondary instruments were not affected by mental loading (figure 6). Subject 11 appears to exhibit some load shedding, primarily on the altimeter and vertical speed indicator.

Fixation Sequences

We are also interested in examining whether pilots develop a scan pattern of patterns during the constant flying maneuver in our experimental paradigm. Assuming that such patterns might exist, it appeared of interest to determine whether they might be altered by the addition of mental loading. The results from one method of studying this question is presented below.

If the dwell times on individual instruments are ignored, an ordered list of instrument fixations may be developed for each pilot for the various mental loading conditions. These lists may be broken up into smaller segments (or sequences) of various lengths for easier analysis. Each different sequence may be considered as a component of the overall scan pattern. One may hypothesize that the sequences which occur most frequently during the maneuver are those of most importance to the pilot and ones which might indicate an ordered scan pattern.

Examination of the results indicated that sequences of four instrument fixations were the longest for which there was a significant amount of repetition during a run, hence sequences of length four were chosen for analysis. The number of times each four instrument sequence occurred during a 10 minute run was obtained as was the total number of sequences of length four in the run. From this data, the percentage of occurrence was calculated for each observed sequence. For example there might be 800 sequences of length four in 10 minutes. If the sequence ATT—DG—ALT—DG occurs 40 times during the run, its percentage of occurrence would be 40/800 X 100 percent = 5 percent. In this fashion, the percentage of occurrence of all length four sequences in the no mental loading case was determined for each pilot. The 10 sequences which occurred most frequently were arbitrarily chosen as indicators of the scan patterns normally used by various pilots. The manner in which the percentage occurrences for these 10 sequences change as a function of mental loading is shown for two subjects in figures 7 and 8. Figure 9
plots the sum of these percentages across mental loading conditions for all subjects. It is important to note that the sequences used as the basis for calculation for all conditions are the ten most frequent for the no mental loading case. Each line beginning at the no mental loading case and ending at the 2-second interval case represents the same sequence.

Several interesting observations may be made by comparing the plots of the skilled pilots with those of the novice subjects. A difference may be seen between the two groups in the percentage of occurrence of the most often used sequences. The first ten sequences used by the skilled pilots comprise over 50 percent of their scan pattern (see figure 9). The usage of these ten sequences is relatively constant with changes in mental loading suggesting that the patterns are not disturbed by the mental loading task. This finding is certainly in keeping with the intuitive development presented in the introduction which suggested that it should be difficult to interfere with a skilled task.

The novice pilots' results differ in several respects from those of the skilled subjects, however. The ten most frequently used sequences in the no loading case occupy much smaller percentages of the total scan than do those of the skilled pilots. This suggests the novices' scans are more random than those of the skilled subjects, even without the imposition of an additional task.

The novice subjects also show a consistent decrease in the percentage
occurrence of the ten sequences as the mental workload is increased. This decrease may be the result of either the equalization of the number of occurrences of each sequence in the run (i.e., a trend to randomization) or a change to a different set of sequences from those used in the no loading case (i.e., a change in strategy). These results support our original hypothesis of a change in the basic scan pattern as mental workload is increased, but indicates the effect is much more evident in pilots of moderate skill.

DISCUSSION

Our results suggest that in a skilled task such as piloting, in which visual scanning plays an important role, the scanning behavior may serve as an indicator of both workload and skill. Before discussing some of the implications of this finding, it is important to note from the outset that the results presented do not seem to support the notion of an accurate, absolute measure of workload. However, a quantitative relative comparison of mental workload under varied conditions does appear to be feasible.

One implication of the results applies to the estimation of workload of some new procedure which may have several possible levels. In many cases, test pilots with superior flying skills are utilized in the estimation or measurement of workload. This often results in equivocal results when comparing alternative procedures, controls, or displays. The present results suggest that different levels of workload may be difficult to measure in such subjects since they appear to be less sensitive to increased difficulty. Our results suggest that pilots of moderate skill are more sensitive to the verbal mental loading task. Thus, if one is concerned with the question of the effect of changing the level of difficulty of some task, then use one step in the evaluation, the use of pilots of intermediate skill at several mental loading levels would seem appropriate since their behavior (visual scanning and performance) will be altered more as a function of the mental loading task than will that of more skilled pilots.

Another possible application may be the assessment of pilot skills. The results have suggested that there is a relationship between the scanning behavior of the pilot and his skill level. The obvious place one might use this technique is in training. One may hypothesize that as a pilot's skills develop, his visual scanning behavior will be less and less affected by non-visual increments in workload. Specifically, it appears that as skill increases, the percentage of long dwells decreases for a particular mental loading level. The scan pattern used during a fixed maneuver is also unaffected by mental loading at higher skill levels. This finding might be utilized in assessing pilots' currency, competency, and level of skill; the technique might be used to pinpoint areas which may require additional training or practice.

The work described here has barely scratched the surface on the issue of prediction of performance via workload measurement. The results suggest that this will be possible in at least some types of situations. In order to examine this matter carefully, performance on both the piloting task and the verbal mental loading task must be closely monitored.

ORIGINAL PAGE IS OF POOR QUALITY
Instrument Scan - Is it an Indicator of the Pilot's Workload?

by

Biomedical Engineering Center for Clinical Instrumentation
M.I.T., Cambridge, Massachusetts

ABSTRACT

This paper describes an investigation of the relationship between an aircraft pilot's visual scanning of instruments and his level of mental activity during a simulated approach and landing. This study is motivated by the increasing concern in several areas of man-machine interaction with the effects of changes in manual control and monitoring procedures on mental workload. This concern is particularly keen with regard to airline pilots, air traffic controllers, power plant operators, and personnel in control of large ocean-going vessels, since the cost of error can be quite high in any of these man-machine systems.

Visual scanning behavior plays an important role in each of these systems, since the operator will typically be required to monitor a number of instruments which display system state variables. In each of the above roles, the human acts as a decision maker, a planner, a manual controller, a monitor, and an event detector. His ability to perform these tasks is generally influenced by their nature, number, and temporal arrangement, by his general physical and psychological state, and by the occurrence of unusual or rare events such as mechanical failures, bad weather conditions, etc.

One may speak of the ability to carry out such tasks in terms of total capacity. Total capacity is a hypothetical limit on tasks which may be performed concurrently and within a certain time period. Under this definition, a person working at 100% of capacity has no resources available to handle additional tasks, while one working at 70% of capacity could be said to have 30% available capacity which might be applied to additional tasks or held in reserve for use in an emergency.

One must be particularly concerned with periods of extremely high or low utilization of capacity, since experience shows that these tend to be the times at which an operator is most prone to error. In the case of high loading, the error(s) may result from inability to accomplish all required tasks within an allotted time period, or failure to detect some item of critical importance (e.g., aircraft altitude several hundred feet lower than expected during an approach). At periods of low loading, on the other hand, errors may result
simply because of a low attention level induced by long periods when little or nothing is happening (e.g. long distance flights over the ocean where the aircraft is controlled by the autopilot and the number of other planes along the airway is low).

Ideally, then, a human operator's job should be designed in such a way as to require an appropriate fraction of the operator's capacity. To accomplish this design objective, however, the designer must have a method at his disposal of estimating the expended capacity under different conditions. While there exists a number of these methods, none is sufficiently benign and non-invasive to be used in the field (for instance, in an airliner's cockpit in flight). Consequently, we have set out to develop an estimator of mental loading based on the operator's visual scan pattern.

In the current work, experiments were conducted in a Terminal Configured Vehicle (TCV) fixed base flight simulator at NASA Langley Research Center. Three NASA test pilots were presented with a piloting task, an arithmetic task designed to vary mental loading, and a side task for calibration of the mental loading task. The pilot lookpoint was obtained by using a highly modified Honeywell oculometer system, and the pilot's eye scan of the instruments was recorded. The piloting task involved flying a curved Microwave Landing System (MLS) approach from a specified waypoint to touchdown. To aid in data analysis, the approach was divided into six segments: downwind, turn to base, base, turn to final, final, and flare. The pilots were aided by a new generation of flight instruments based on CRT displays which were installed in a simulator. These were an Electronic Attitude Direction Indicator (EADI) and an Electronic Horizontal Situation Indicator (EHSI) in place of the conventional flight director and horizontal situation indicator. The EADI provides conventional flight director information such as localizer and glide slope deviations, and pitch and roll attitudes. It also provides additional features including flight path angle, flight path acceleration, radar altitude, and a dynamic perspective drawing of the runway. The EHSI is a moving-map display with ownship at the center. During the MLS approach, the curved MLS glidepath is drawn and the pilot may use various optional features to allow navigation. Features include trend predictor vectors to show aircraft position up to 90 seconds in the future and display of all other aircraft (traffic) in the approach pattern. For further discussion of these displays, see Harris and Mixon (1979).

The mental loading task was chosen so as not to interfere with the visual scanning of the pilot while providing constant loading during the approach. This was accomplished by having the pilots respond verbally to a series of evenly spaced three-number sequences. The pilot was told that he must respond to each three-number sequence by saying either "plus" or "minus" according to the following algorithm: first number largest, second number smallest = "plus"; first number smallest, last number largest = "plus"; otherwise = "minus". The numbers were recorded at twenty second and ten second intervals. These intervals had been determined empirically to vary mental loading under a similar piloting task.

The workload measuring side task employed two lights, one mounted above the other, placed just outside the pilot's peripheral view above the instrument panel. The lights came on at random intervals between one and three seconds and remained on for one second. The pilot was told to turn the lights off by using a three-position rocker switch on the control grip (moving the switch up turned the upper light off, down turned the lower light off). This was done only when the pilot had time left from performing the primary task of flying
the airplane. Thus the number of correct responses to the lights gave a measure of the residual capacity of the pilot from which a workload index could be calculated.

The experimental conditions were arranged in a 2 x 2 x 3 factorial design. The conditions were the presence or absence of traffic (other airplanes in the same approach pattern) on the pilot's EHSI display, presence or absence of the side task lights, and mental loading task (no numbers, three number sequences at twenty second or ten second intervals). Two replications were obtained for each pilot. Of the twelve runs per replication, only the six involving no light-cancelling side task were used to study the scanning behavior, since the presence of the side task lights would alter this behavior.

Results of the side task showed a definite increase in workload when the arithmetic task was introduced. The x-y plots of pilot lookpoint for each segment of the approach also show substantial qualitative differences between the different levels of loading. The three instruments used most by the pilot in the scan are the EADI, EHSI and the air-speed indicator. The largest number of transitions were within the EADI, while the next largest were between the EADI and the EHSI, followed by the airspeed and other instruments. The detailed scanning within the EADI is of particular importance. The display is used almost exclusively during final approach and flare, those segments when workload is usually judged subjectively to be the highest.

A computer algorithm has been developed to obtain the first-order, discrete-state, discrete-transition, Markov model for each pilot's scanning pattern. It is assumed that workload is constant within each of the six approach segments since the piloting tasks are essentially constant over each segment. This allows comparison of the instrument transition matrices for each segment with those obtained under different loading conditions. The relationship between visual scanning and workload is given by the change in the elements of these matrices as loading varies. Higher-order Markov models may also be used to provide a more accurate description of the processes taking place.

The assistance of R.L. Harris, Sr., J.Keyser, L.Person, and R.Yenni, all of NASA Langley, is gratefully acknowledged. This work was supported by NASA Cooperative Agreement NCCI-23.

References