(NASA-CR-173033) BEHAVIORAL INDICATORS OF PILOT WORKLOAD (Columbia Univ.) 12 p HC A02/MF A01 CSCL 05I

N83-34580

Unclas
G3/53 15127

DRA

Psychophysics Laboratory

Columbia University
in the City of New York

IN LVMINE TVO VIDEBIMVS LVMEN

This report and the research herein described has been prepared with the partial support of the Office of Naval Research, the U.S. Army Medical Research and Development Command, and other agencies of the United States Government. Inquiries concerning the contents of this report should be made to Professor Eugene Galanter, Department of Psychology, Columbia University, New York, New York 10027.

1 May 1983

Columbia University
Psychophysics Laboratory
324 Schermerhorn Hall
New York, NY 10027

Technical Report PPL 83/4

Behavioral Indicators of Pilot Workload

Eugene Galanter and Julian Hochberg

Presented at the Second Symposium on Aviation Psychology

Columbus Ohio, April 26, 1983

This research was supported in part by Cooperative Agreement NCC 1-5 -- THE INFLUENCE OF AUDIO INPUTS ON SIMULULATED FLIGHT OPERATIONS -- between the NASA Langley Research Center and Columbia University. The technical officer for this program is Dr. Randall L. Harris Sr.

Behavioral Indicators of Pilot Workload

Eugene Galanter and Julian Hochberg

Columbia University

ABSTRACT

Using a technique that requires a subject to consult an imagined or remembered spatial array while peforming a visual task, we show that there is a reliable reduction in the number of directed eye movements that are available for the acquisition of visual information.

In earlier experiments*, eye movements recorded during a <u>primary task</u> (multiple dial readings) dropped precipitously whenever a sporadic <u>probe task</u> required subjects to report the orientation of the corners of a memorized figure, showing greater demands with more demanding figures.

In presently reported research, these results were extended and replicated with a primary task based on high-fidelity video simulations of runway approach to landing, auditory probe instructions, and simple video recording of gaze direction suitable for cockpit use.** Further replication in a General Aviation cockpit simulation (GAT II at NASA Langley)*** suggests that the measurement technique can be applied to pilot workload assessment in actual flight.

INTRODUCTION

For the last 5 years, we have been developing and testing a method to provide an online, (covert and non-invasive) measure of cognitive load. (Hochberg and Galanter 1980) In this paper, we describe a version of this method that relies on perturbations of normal eye movements that a pilot must execute around known places in the environment on an esssentially continual basis. The method appears to be sound and applicable to a wide variety of situations, and should now be tested under such field conditions.

The logic of the method, and a summary of the assumptions that underlie its applications, is developed in the first three figures. In Figure 1, the cognitive loads imposed by each of three arbitirary tasks (T1, T2, T3) are shown below a horizontal line, midway up the ordinate, that symbolizes the operator's capacity limit for effective cognitive function. Note that none of the tasks, considered individually, exceeds that limit.

Although they do not universally do so, concurrent task loads normally pool their effects in some fashion. The conditions for such pooling, and the function according to which the pooling occurs, are not known with any precision at present; indeed, one motive for finding a measure for cognitive

ORIGINAL PAGE IS
OF POOR QUALITY

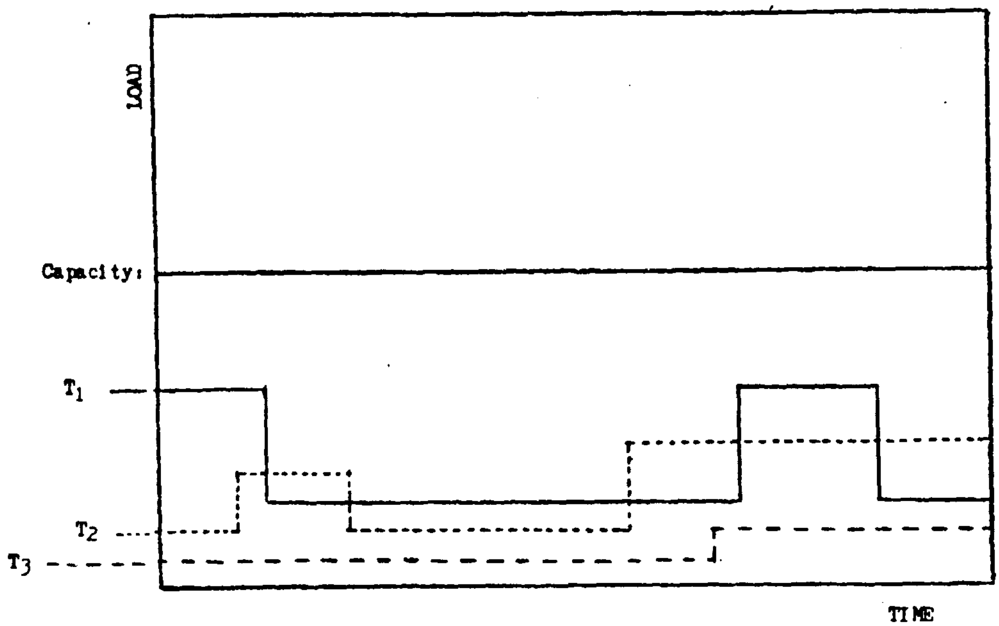


FIG. 1

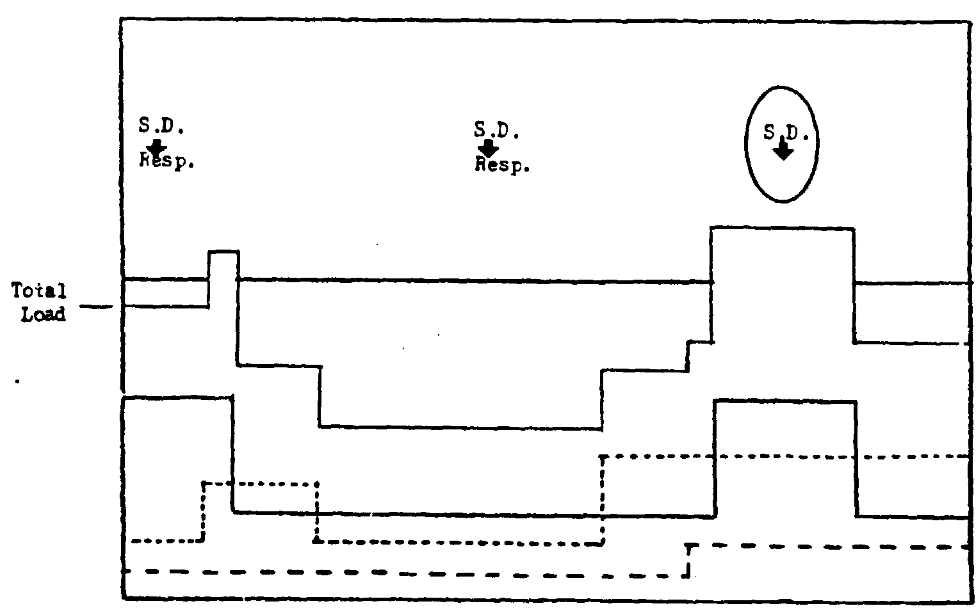


FIG. 2

load is to explore these questions. In Figure 2, we represent the joint effects of the three tasks as an unweighted sum. We do this in order to make the point as simply as possible, with no intention of implying that such linear summation is the general case.

Note that although the tasks individually do not breach the subject's capacity limits, they do so when performed in parallel. Exceeding the capacity limits need not have any overt consequences; indeed, when that occurs in an interval in which the situation makes no demand upon the subject (as in the first such overload in the figure) the overload may not be evident even to the subject. But when a situational demand (S. D., in the figure) occurs during an overload interval, as in the second occurrence, the subject's response should reveal tha fact in some way. Such reactions to the cognitive overload may range from unnoticed response delay, through decision errors, to a collision with the ground.

According to this analysis, we need one of two measures in order to evaluate the demands on the operator: (a) Ideally we would want a quantitative but nonintrusive measure of the total cognitive load, tracking its curve regardless of situational demands. Unfortunately, we do not know of any method that provides such a measure. (b) Alternatively, we need a method which gives a continuous measurable indication of when the capacity limits have been breached, regardless of whether or not an external situational demand has been presented. Given such a measure, we could then determine the total load at any point in time by the addition of some probe task. This task should be of known but controllable load that can be varied in magnitude. In this way we could manipulate the load imposed by the probe task which in turn brings the total load to the capacity limits. This would allow us to titrate the load provided by the nonprobe tasks, and thereby assess their effective magnitude.

In Figure 3, such an indicator is provided by the horizontal line shown immediately below the capacity limit, which is breached whenever the total load exceeds the limit.

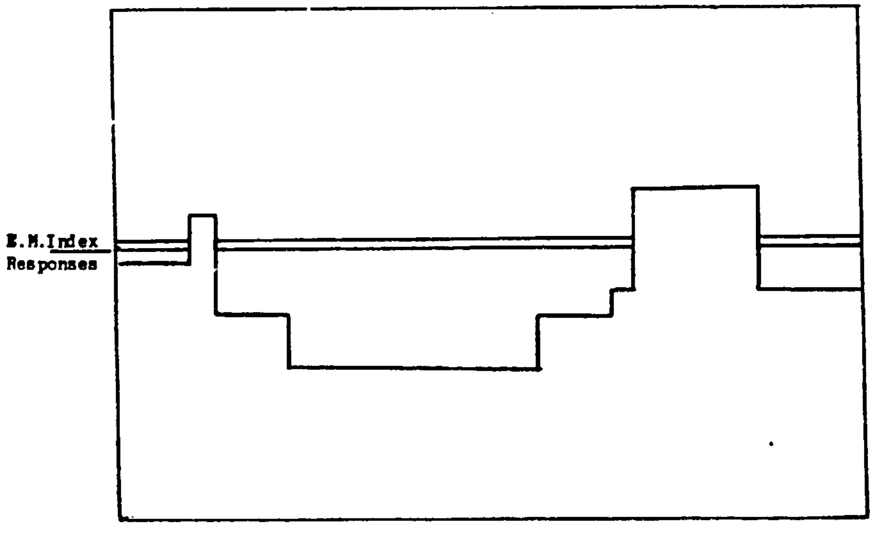


FIG. 3

The indicator task that we have been using is provided by the continuing need to update information from several fixed points in the pilot's environment (e.g., specific instruments in the cockpit which change at some known rate, or the point of projected touchdown on the landing strip ahead, etc.). Because the places at which the pilot must direct his eyes are fixed, the rate at which his direction of gaze shifts between them is (at least in principle) measurable with quite inobtrusive methods.

We have reason to believe that the updating of information, and indeed the execution of eye movements themselves, are suspended when cognitive load is exceeded. Using this supension of eye movement as a measure should provide a usable index. In any case, because the indicator task can be one in which the updating behavior is "deferrable" -- i.e., the pilot must only obtain the information from each source within some time window determined by the rate of change of the display and by the complex of task demands, rather than on some fixed schedule — the level at which the indicator task is disrupted could be held at some point below the operator's capacity limits.

METHOD

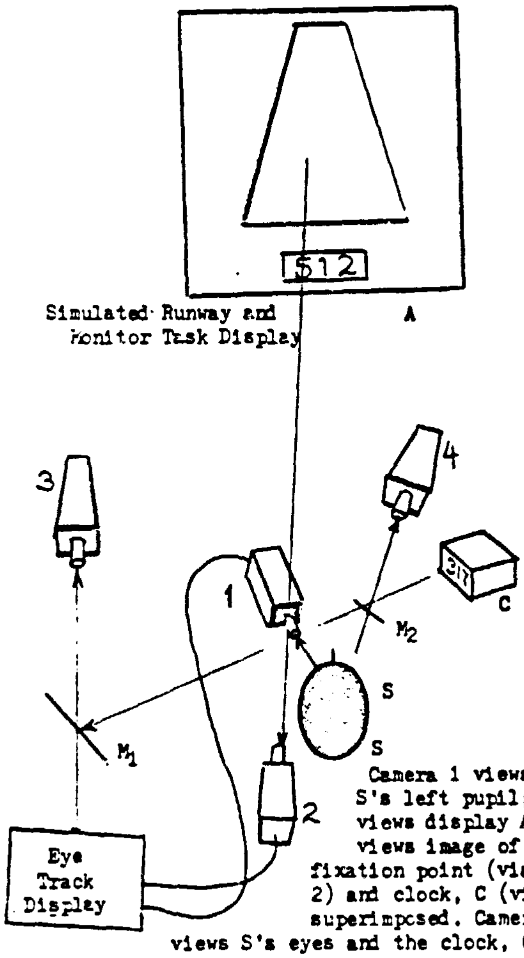
Given a situation, then, in which the ongoing tasks, including the indicator task which is revealed in the continual eye movement behavior, do not exceed the capacity limit, it should be possible to impose a probe task that raises the total load. The level of interference with the indicator task and the eye movement behavior that task generates, can be controlled by the experimenter. This permits us to test the stability of the indicator task, and therefore the reliability of the probe.

That is the paradigm for the research that we have been doing. In what follows, we describe the essential features and results of several studies. Some of these experiments were done in collaboration with Mary Peterson and Dale Klopfer, one was done in collaboration with Richard Popper, and one was substantially assisted by Nancy Haber.

Apparatus

We will make the method more concrete by illustrating the setup used in the last of our series of experiments. In this experiment, both of the systems of gaze measurement that we used in the prior series were employed simultanecusly. Since one method (video recording) is at least an order of magnitude less expensive and instrumentally simpler, we wanted to determine whether data provided by this method was as useful as that generated by a full-fledged eye-tracker.

In Figure 4 the subject's head is shown at S; the display at which he or she looks is a video screen at A, on which is shown a filmed segment of an approach to an airport runway, with a continuously changing set of numbers superimposed at the bottom of the screen. An infrared video camera (1) is aimed at the subject's left eye; the output of the camera is fed to a G+W (Whittaker) processor which (after suitable calibration) superimposes the gaze direction on a view of display A, as picked up by the video scene camera (2). Video camera (3) records on videotape both the eyetrack display (i.e., the output of cameras 1 and 2), and, in addition, the view of an LED clock (c) which is reflected in a half-silvered mirror (M1).



Camera 1 views S's left pupil; 2 views display A; 3 views image of A with fixation point (via 1 and 2) and clock, C (via M_1) superimposed. Camera 4 views S's eyes and the clock, C, via M_2.

FIG. 4

The inexpensive method of measuring gaze direction was based on filming the subject's eyes by video camera (4), which was aimed so as to photograph both the subject's eyes and the image of the clock (c) as reflected in beamsplitter M2. Camera (4) was really mounted behind the subject, with the appropriate configuration of optical paths, but is represented here as being in front, for ease of illustration. The output of camera (4) was also recorded on videotape. The two sets of videotape recordings could be compared by using the image of the clock which appeared in both tapes and served as a calibrator.

Procedure

Throughout this series of experiments, the subjects had two kinds of tasks: a set of indicator tasks and the probe task.

In our earlier experiments, the indicator task required the subject to monitor two or more dials with similar response demands. In these experiments we established the parameters of stimulus variation that permitted some degree of deferral of information acquisition. In all of these monitoring tasks, the target events occurred in a pseudorandom sequence, and the average rates of change, and the payoff matrices, were such as to permit some deferral of updating most of the time.

The indicator tasks (or visual monitoring tasks) in the experiments shown in Figure 4 required that the gaze be continually switched to two or three different places. In these experiments the subject was required to monitor the runway and to press a (index finger) trigger as rapidly as possible every time the center of the runway crossed an arrow marked on the screen. The runway drifted back and forth across the screen in a continuous but pseudorandom fashion. The subject was also required to press a (thumb) trigger whenever the number 500 appeared in the display at the bottom of the screen. These numbers fluctuated around the value of 500 in a staircase pattern of pseudorandom length and direction.

Because both stimuli changed in a continuous and relatively predictable fashion, subject's could anticipate critical events with some degree of accuracy. Payoffs were adjusted to ensure that successful performance could be obtained even though there was some intermittancy of data acquisition.

The probe task was essentially the same in all experiments. Subjects first practiced naming each of the four quadrants (I, II, III, and IV) at the bottom of Figure 5. These names are the standard nomenclature used in mathematics for the quadrants of the Cartesian coordinate system. They were then shown that the corners of outline letter shapes (shown at the top of Figure 5) could be interpreted as quadrants. By use of the clockwise numbering scheme shown in the top two patterns of Figure 5, subjects could be asked the quadrant identity of any corner. The figures could vary in the number of corners they had, whether they were presented in normal form, or rotated as in the "lazy" C in Figure 5. The subject could also be asked about corners at various removes from the starting point.

On an irregular schedule during the experiments themselves, one or another letter pattern was named through earphones, a corner of that pattern was identified by number, and the subject was told to respond as quickly as possible by identifing the quadrant name of that corner. The visual imagery problem presented to the subject by this task thus varied in its "depth," i.e., in the number of corners that the subject had to count off in his "mind's eye" before reaching the target corner.

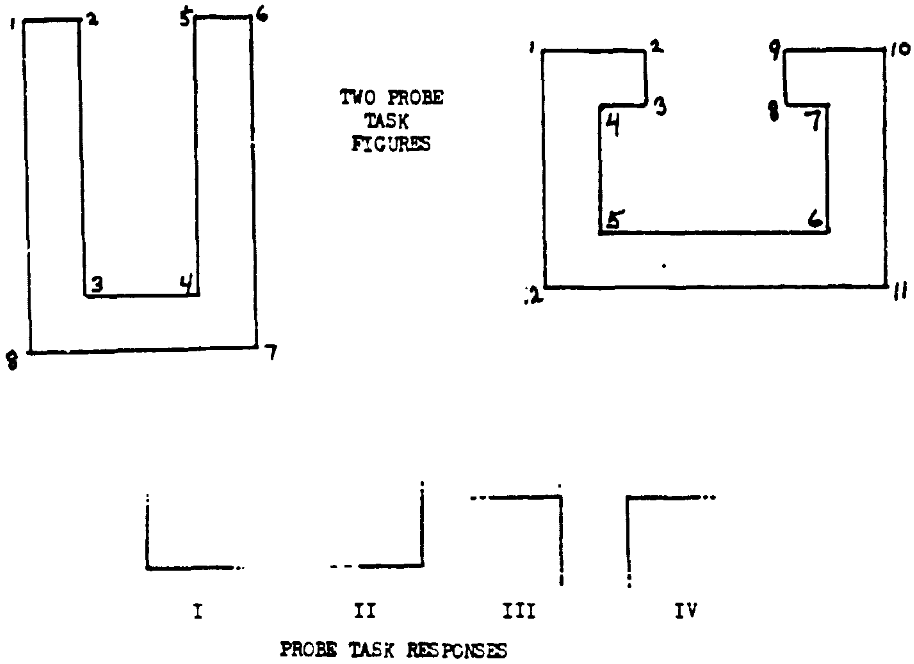


Figure 5

RESULTS

If the argument developed in connection with the first three figures is valid, and if the total load consisting of the probe task added to the indicator task presses the limits of the subjects' capacities, we would expect the eye movement rate to drop during the execution of the probe tasks. The upper panel in Figure 6 is a sketch of a gaze direction record above, and the time course of an aperiodic presentation of the probe task below, where S signifies the auditory presentation of the task and R signifies the subject's response.

The lower panel of Figure 6 summarizes the results of the major experiment in our series (done in collaboration with Klopfer and Peterson), in which the indicator task involved monitoring the changing numbers in three dials, and gaze direction was measured by the eye tracker and recorded by camera (3) in Figure 4. In using the eye tracker, gaze direction had been sampled at 250 msec intervals, and the reciprocal of the duration of any fixation was taken as a measure of eye movement frequency and attributed to each 250 msec bin within that duration.

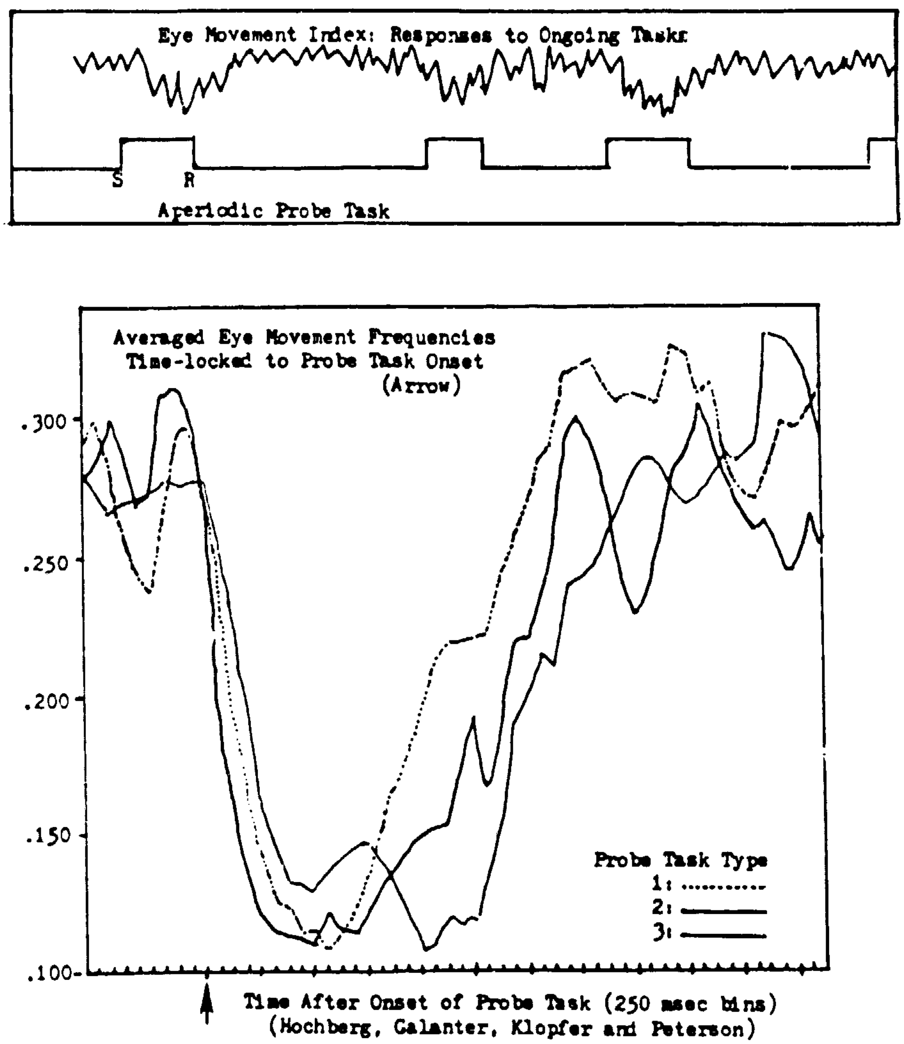


Figure 6

The lower panel in Figure 6 shows the actual course of such averaged eye movement frequencies obtained in that experiment, time-locked to each onset of the probe task (the onset time is shown by the arrow on the abscissa). The ordinate reads in an equivalent of the reciprocal of dwell time. Thus, before and after the probe tasks, the average duration of fixations was (roughly) three seconds. During the suppression the dwell dropped to more than eight seconds. The abscissa reads in 250 msec bins. The probe tasks were divided into three types, shallow (1), medium (2), and deep (3). Note that the duration of the interval during which indiator task frequency was depressed was clearly a function of the depth of the probe task.

Similar results were obtained in an experiment (assisted by Nancy Haber) using the runway-and-number version of the indicator task, as illustrated in Figure 4, using judges' estimates of gaze-direction changes, as recorded by the simple video setup indicated by camera (4) in that figure. We have concluded that such free-camera recording of gaze direction may have virtues for actual cockpit use, and a preliminary analysis of the last experiment in the series (in which both measures of gaze direction are combined, as shown in Figure 4) shows that the two measures yield similar results. (Analysis of the data in this last experiment is not fully completed -- free-camera recording is easy on the subject but hard on the scorer.)

Because further replication of the scoring procedure has been done but has not yet been completely analyzed, this last experiment must be treated with some caution, but the fact that eye movement frequency drops during execution of the probe task, as shown in Figure 5, is robust in any case.

DISCUSSION

The use of measurable behavior of an indicator task, taken in conjunction with a probe task, would seem to be an extremely useful tool for measuring cognitive load in actual flight. It may therefore serve to measure the skill of the pilot as a result of different kinds of training, and as a function of different kinds of instrumentation. This workload measure may also be useful to assess the effects of different operating conditions and procedures.

There remains the question of whether the indicator task behavior was depressed by the probe task because the former involved eye movements and the latter involved visual imagery. If this is the case, it may make the measure specific to one class of behavior and one kind of load. On the other hand, increases in cognitive load of any kind at all may be reflected similarly in the probe task behavior (Tole, Stephens, Harris and Ephrath, 1981). That is a matter for future laboratory research. Before that is undertaken, however, it would probably be desirable to develop and test this method of measuring cognitive load in actual flight conditions in order to learn whether the technique will actually work for the intended purposes. There is little point in refining the technique if it cannot be applied where it is needed.

REFERENCE NOTES

* In collaboration with Dale Klopfer and Mary Peterson
** Assisted by Nancy Haber
*** In collaboration with Richard Popper

REFERENCES

Hochberg, J. and Galanter, E. Acoustically induced visualization inturrupts visual information acquisition, PPL 80/5, Psychophysics Laboratory, Columbia University, New York, 1980.

Tole, J. R., Stephens, A. T., Harris, R. L., & Ephrath, A. E. Visual scanning behavior and pilot workload, NASA Technical Paper VSWLP.SC, 1/12/81.