FINAL SUMMARY REPORT FOR
PHASE I SBIR PROJECT #F33615-88-C-0545:

# UNOBTRUSIVE REAL-TIME MONITORING OF PILOT MENTAL STATUS: DEVELOPMENT OF A TEST-BED SYSTEM

Submitted to:

Armstrong Aerospace Medical Research Laboratory
Human Systems Division
Attn: AAMRL/HEG (Dr. Glenn F. Wilson)
Contract No.: F33615-88-C-0545
Wright-Patterson AFB, Ohio 45433-6573



Submitted by:

Dr. Richard L. Horst
Dan T. Smith
John K. Nebel

ARD Corporation
9151 Rumsey Rd.
Columbia, Md. 21045
(301-596-5845)

June, 1990

USAF/803-011-90

**ACKNOWLEDGMENTS**

The work reported here was conducted under an ARD Corporation, Phase I, Small Business Innovative Research contract (F33615-88-C-0545) from the U. S. Air Force Human Systems Division. Dr. Glenn Wilson of the Armstrong Aeromedical Research Laboratory, Wright-Patterson Air Force Base, was the Contracting Officer's Technical Representative. We are grateful to the Rotorcraft Human Factors Group at NASA Ames Research Center for supplying us with the Window/PANES task software used here, as well as guidance on how to modify it for our purposes, to the Northrop Corporation for the use of their Delta-Biometrics Vagal Tone monitor, to Dr. Stephen Porges of Delta-Biometrics for the loan of an ECG amplifier, to Dr. David Paskewitz of the University of Maryland for the loan of the polygraph used for EEG/EOG amplification, and to Susan Graham and Harry Wujek of ARD Corporation who prepared the graphics for this manuscript. A video-taped presentation and demonstration of the subject test-bed which was also delivered to the Air Force under the present contract. Copies of this tape in VHS format are available upon request from Man-Made Systems Corporation, 4020 Arjay Circle, Ellicott City, Md., 21043, (301-596-4915).

## ABSTRACT

The real-time determination of pilot mental and physical status is a critical feature of the workload monitoring and "Mindware" subsystems that have been envisioned for future jet aircraft. Recent laboratory and simulator studies, using retrospective data analyses, have suggested the value of various behavioral and physiological indices for reflecting task performance. The purpose of the present work was to develop software algorithms to derive some of these measures of interest in real-time and to develop a test-bed in which to explore the efficacy of these measures for inferring operationally relevant changes in pilot status.

The present project demonstrated the feasibility and usefulness of this approach. Data processing algorithms were developed for characterizing and integrating physiological indices based on heart-rate and heart-rate variability (vagal tone), eye blinks, and single-trial, scalp-recorded event-related potentials. These physiological measures were obtained concurrently with behavioral measures as subjects performed a PC-based, aviation simulation task (Window/PANES). The data processing algorithms were implemented in a distributed processing configuration, using multiple personal computers, with the derived measures being integrated by a "Decision-Maker" processor. This multi-processor test-bed was demonstrated to work in near real-time (with lags of five to ten seconds), and attained encouraging levels of accuracy in characterizing the physiological phenomena of interest. Also encouraging were preliminary efforts to define decision rules, customized for individual subjects, that reflected a sensitivity of the measures derived in "real-time" to manipulations of task difficulty.

The present test-bed should prove useful in supporting future studies of dynamic decision-aiding and task partitioning between the pilot and on-board automation. It also lays the groundwork for the future development of an integrated, real-time performance monitoring workstation that would combine the functionality of the individual data processors and Decision-Maker processor used in the present feasibility demonstration.

## TABLE OF CONTENTS

## LIST OF FIGURES

# 1.0 BACKGROUND

## 1.1 Introduction

The design of present-day and near-term future fighter aircraft have advanced to the point that they are becoming increasingly difficult for a human pilot to fly. The maneuverability and sensor technology that is being built into these aircraft provide for truly remarkable advances in mission capabilities for terrain following ingress to and egress from an area of hostilities and the ability to take evasive action when challenged. However, the aircraft can now perform maneuvers that cause such high-G loads that the pilot is at risk for losing consciousness, and when in control, the pilot is presented with such a bewildering array of information to be processed and attended to simultaneously that his performance may deteriorate due to cognitive overload or stress-induced mental fatigue.

In response to this worsening design problem, a variety of innovative man-machine interface technologies are being considered. These approaches, which attempt to take advantage of emerging capabilities in micro-processor based display technology and artificial intelligence, include such areas as voice interfaces, virtual displays (not only in the visual modality but also in the auditory and tactile modalities), and automated "pilot associate" aids for the aircrew. A number of these emerging technologies have been captured in the visionary ideas for a Super-Cockpit, which will provide the focus for much of the man-machine interface research in the next decade.

As these new interface possibilities mature, however, a host of new design issues arise, pertaining to the most effective role of the human operator in a system with ever more automated subsystems. Under what circumstances and to what extent should automated subsystems be used to aid the human operator? What should be the default conditions? In other words, do we rely on the human operator to call for help, do we have the automated aids available but non-intrusive, in case the human decision-maker feels they are irrelevant to a given situation, or do we design for automated intervention, with the capability for operator override?

It is a continuing truism that a trained human operator, functioning normally, has certain mental capabilities for pattern recognition and decision-making that are not yet achievable by artificial intelligence. Therefore, given sufficient cognitive capacity on the part of the operator, many "higher level," supervisory functions in the control of complex systems are best left to the human. However, on-board automation capabilities have developed to the point that, if the human operator is debilitated either physically or mentally, due to cognitive overload or fatigue, it is preferable for certain functions to be taken over by automated subsystems. Therefore, the picture that emerges for future man-machine interfaces in many systems, including advanced aircraft, is one of tasks being dynamically traded off between man and machine, depending on the moment-to-moment demands of the environment and the corresponding functionality of the human operator.

## 1.2 The Need for Measures of Mental Workload.

In order for such dynamic control scenarios to be realized, we must develop reliable indices of operator mental workload. Moreover, if these workload measures are to be used operationally, whether they are a primary input to an onboard performance monitoring systems that dynamically allocates tasks or whether they are used to determine retroactively whether or not the human-in-the-loop made reasonable choices, moment-to-moment, about his ability to function without automated assistance, these workload determinations must occur in real-time.

For the time being, it is sufficient to define mental workload loosely and operationally, as the sum total of information-processing burdens, whether self-imposed or environmentally imposed, which occupy mental capacity and thus prevent certain additional information from being processed. Note that although operator effectiveness is ultimately defined in terms of behavioral output, it is valuable to conceptualize mental workload as a construct in addition to focusing just on observable behavior. Behavioral performance may deteriorate for a wide variety of reasons, for example due either to cognitive overload or boredom, so the workload construct has some diagnostic, and hopefully prescriptive value. On the other hand, most tasks allow the human operator to function with some spare capacity such that, to some extent, increased task demands can be met with increased effort in order to maintain behavioral output at a relatively constant level. For this reason, it is valuable to conceptualize workload indices that are not based solely on behavior, in order to predict susceptibility to impending deterioration in performance.

### 1.3 Problems in Assessing Mental Status in Operational Settings.

The development of mental workload measures has been a particularly fruitful endeavor over the last few years (see reviews by Hart, 1986; Gopher & Donchin, 1986; Moray, 1979), but many of the measures that have proven useful for basic research in laboratory or simulator settings are not usable in the operational environment. It is self-evident that real-time workload measures implemented in operational systems must be unobtrusive, or at least obtainable without further burdening the operator with information processing or response demands; otherwise, they defeat their purpose.

It follows from the above discussion that one can not evaluate workload solely from observing an operator's behavior on his primary task. Performance on a secondary task has been an often used approach for measuring the workload of a primary task. However, with this approach it is difficult to ensure that the operator always gives mental priority to the primary task, the results may be of questionable validity if used to generalize to situations in which the primary task is performed alone, and incompatibilities between the behavioral responses required by the two tasks may make it difficult to draw inferences about the demands placed on perceptual or decision-making processes. Moreover, the sort of contrived secondary tasks that have often been used in laboratory studies are clearly not acceptable in operational settings, so secondary task measures must be found among the activities that the operator is doing anyway.

Simply asking the operator for subjective ratings of his perceived state is also fraught with difficulties. The operator may not realize that his environmentally-defined workload is high when in fact it is, such subjective ratings tend to be unreliable when administered in operational settings while the operator is simultaneously trying to maintain task performance, and the mere act of completing the rating itself, of course, constitutes an additional task burden on the operator.

Another possibility is to monitor psychophysiological indices from the operator. As discussed in more detail later, such indices as heart-rate variability extracted from the electrocardiogram (EKG), blink frequency and duration extracted from the electrooculogram (EOG), and event-related potential (ERP) amplitude and latency extracted from the scalp-recorded electroencephalogram (EEG) have all been related to cognitive processing and selective attention (see e.g., the April, 1987 special issue of Human Factors). All of these indices can be recorded non-invasively and relatively unobtrusively, with miniaturized instrumentation such as the Solid-State Physiological Instrumentation Data Recorder (SSPIDR) developed by Systems Research Laboratories for the Navy, the Vitalog Monitoring System developed by Vitalog Corporation, and the DelMar Avionics data recorder. In addition, progress is being made in detecting and correcting data for the numerous electrical artifacts which can impinge on recordings in simulator or operational settings. These physiological measures are not, however, routinely derived in real-time. Furthermore, the absolute levels of physiological indices do not uniquely indicate mental states (see e.g. Johnson, 1980; Zacharias, 1980), the often-cited example of this being that similar EEG activation occurs during attention to a task and upon entering the deepest stage of sleep.

## 1.4 The Present Approach.

It is therefore apparent that no single index of workload will prove sufficient for use in the operational setting. However, an approach that offers a rule-based interpretation of an appropriate constellation of behavioral and physiological indices appears promising. The approach implemented here involved the simultaneous recording and processing of a battery of physiological and behavioral indices. The focus was on changes in these indices with changing task demands, rather than a focus on absolute levels of these indices. These changes can then be subjected to an intelligent, rule-based interpretation which takes into account their time history as well as the inter-relationships among the various derived measures. The outputs of this interpretation are inferences about mental workload and predictors of impending deficits in operator performance in controlling the system and responding to environmental challenges.

The goal of the work reported here was, therefore, to demonstrate the feasibility of thus monitoring, in real-time, the mental status of the operator in a multi-task man-machine control system, using a combination of behavioral and physiological measures. In order to do this, a test-bed was developed within which multiple behavioral and physiological measures can be concurrently obtained and processed in real-time, as an experimental subject performs a simulated aviation task. The scope of this project obviously did not allow a major system development effort. Thus the present test-bed was configured using existing hardware -- a number of IBM-compatible personal computers (PCs) with commercially available plug-in boards to accomplish analog-to-digital (A-to-D) conversion, digital input/output (I/O), and serial communications. The focus was on the development of real-time algorithms for processing the physiological indices of interest in order to obtain useful derived measures in near real-time, and on the development of software to integrate these measures in a way that would allow the system supervisor (i.e., the experimenter) to interactively apply decision rules to these derived measures. The processing algorithms and decision rules that are developed in the context of this test-bed, as well as the distributed processing design philosophy, can later be implemented in a miniaturized "black-box" that could be integrated into operational settings and used to make inferences that will support the real-time measurement of human performance and dynamic task allocation between man and machine.

## 2.0 RESEARCH OBJECTIVES AND GENERAL APPROACH

The specific research objectives of the present project were as follows:

- o Develop real-time versions of several existing data analysis algorithms (for quantifying eye blinks from EOG and cognitive event-related potentials from EEG).
- o Adapt a low fidelity aviation simulation for the present test-bed.
- o Develop "decision-making" software to poll and integrate the various behavioral and physiological indices being computed in real-time.
- o Construct a working test-bed configuration using existing hardware and system software.
- o Collect preliminary data to validate the usefulness of this test-bed and to derive "first-cut" decision rules for delineating operator "mental status".
- o Formulate plans for future uses and development of this technology.

Initial validation of this test-bed involved an examination of the sensitivity of the real-time algorithms to manipulations of the physiological signals, both simulated and actual. Then preliminary investigations were conducted to demonstrate the usefulness of the test-bed approach. Appropriate parameters for the pattern recognition algorithms (EOG, ERPs, and "decision rules") were determined by conventional analyses of the data collected as a subject performed the aviation task during a "training set" session. These parameters were then applied in real-time to the data collected from the same individual during a subsequent "test set" session. The objective during the "test sets" was to determine how accurately known task manipulations of cognitive workload could be inferred from combinations of the derived real-time measures.

# 3.0 DESCRIPTION OF THE TEST-BED

## 3.1 Overview of the Test-Bed

Figure 1 provides a schematic of the multi-processor test-bed that was configured for the present development effort. The hardware components for each subsystem, which in most cases was a separate IBM-compatible PC, are shown in the boxes. The functions of each subsystem are shown on the left side of the figure and the derived measures that are transmitted by each subsystem to the "Decision-Maker" are shown on the right. The directions in which information flows through the test-bed are indicated by the arrows. A schematic of the physical connections involved in this configuration is shown in Figure 2.

A PC-based, low-fidelity aviation simulation was used as the task environment. The tasks involved here included psychomotor tracking, choice reaction time to occasional transient stimuli, and monitoring the level of gauges which indicated system status. Task difficulty and, by inference, mental workload, was manipulated systematically, as subjects performed in this multiple-task environment. The aviation simulation yielded behavioral measures of primary task performance (root-mean squared tracking error) and secondary task performance (choice reaction-time and accuracy of responses to the transient stimuli and to the occurrence of abnormal conditions indicated by the gauges).

Physiological measures included heart rate and vagal tone derived from the ECG, the frequency of occurrence, duration, and timing of eye blinks derived from the EOG, and the amplitude and latency of several endogenous components of the scalp-recorded ERPs derived from the EEG. A distributed processing approach was implemented whereby a separate PC processed each type of physiological signal. The resulting derived measures were conveyed, by serial communications, from these individual processors to a "Decision-Maker" processor. The Decision-Maker polled and stored these incoming data and implemented some simple pattern recognition algorithms to allow the triggering of "cautions and warnings" when certain measures exceeded pre-selected set-points. The Decision-Maker also provided an interface for the experimenter

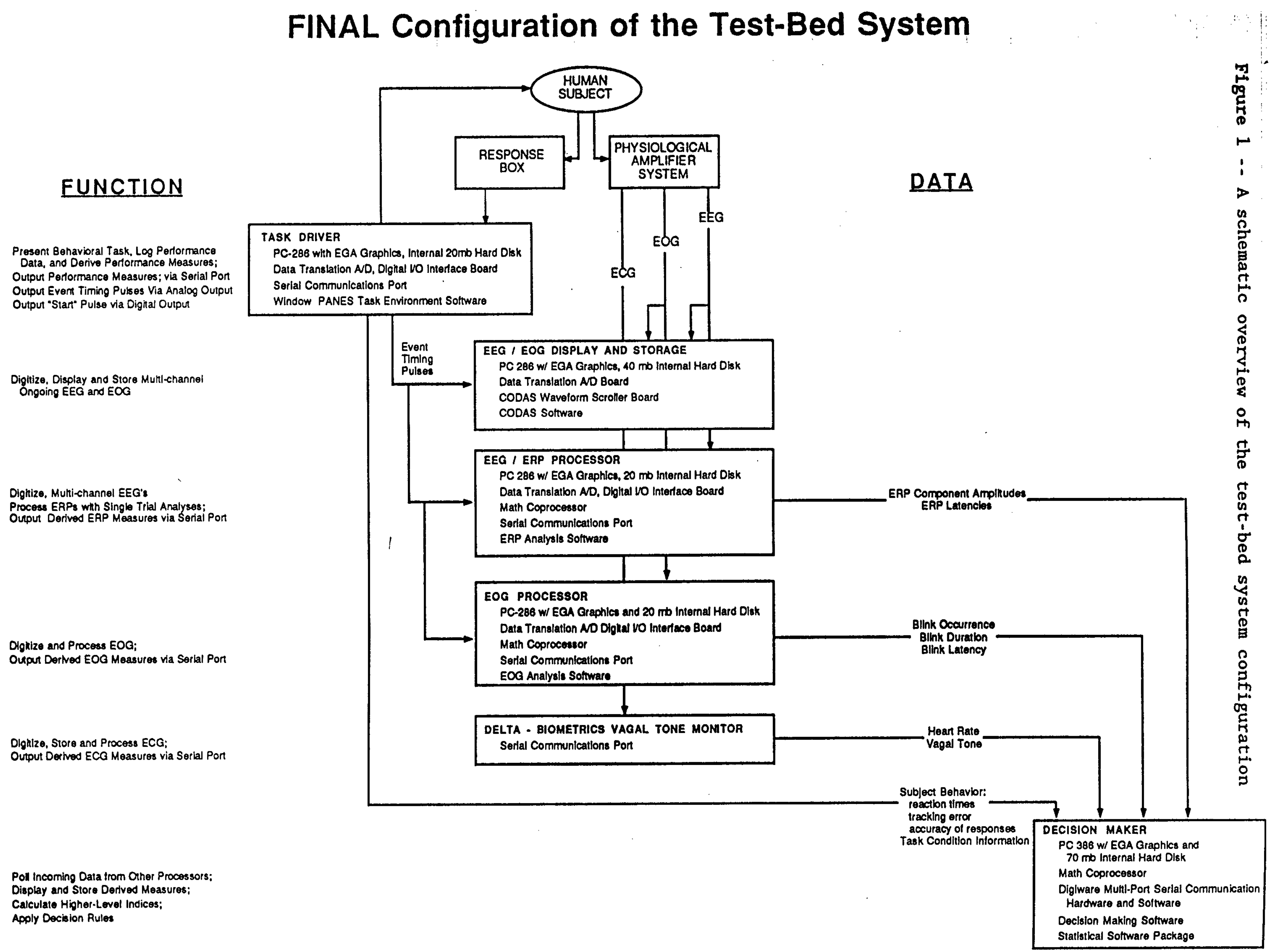
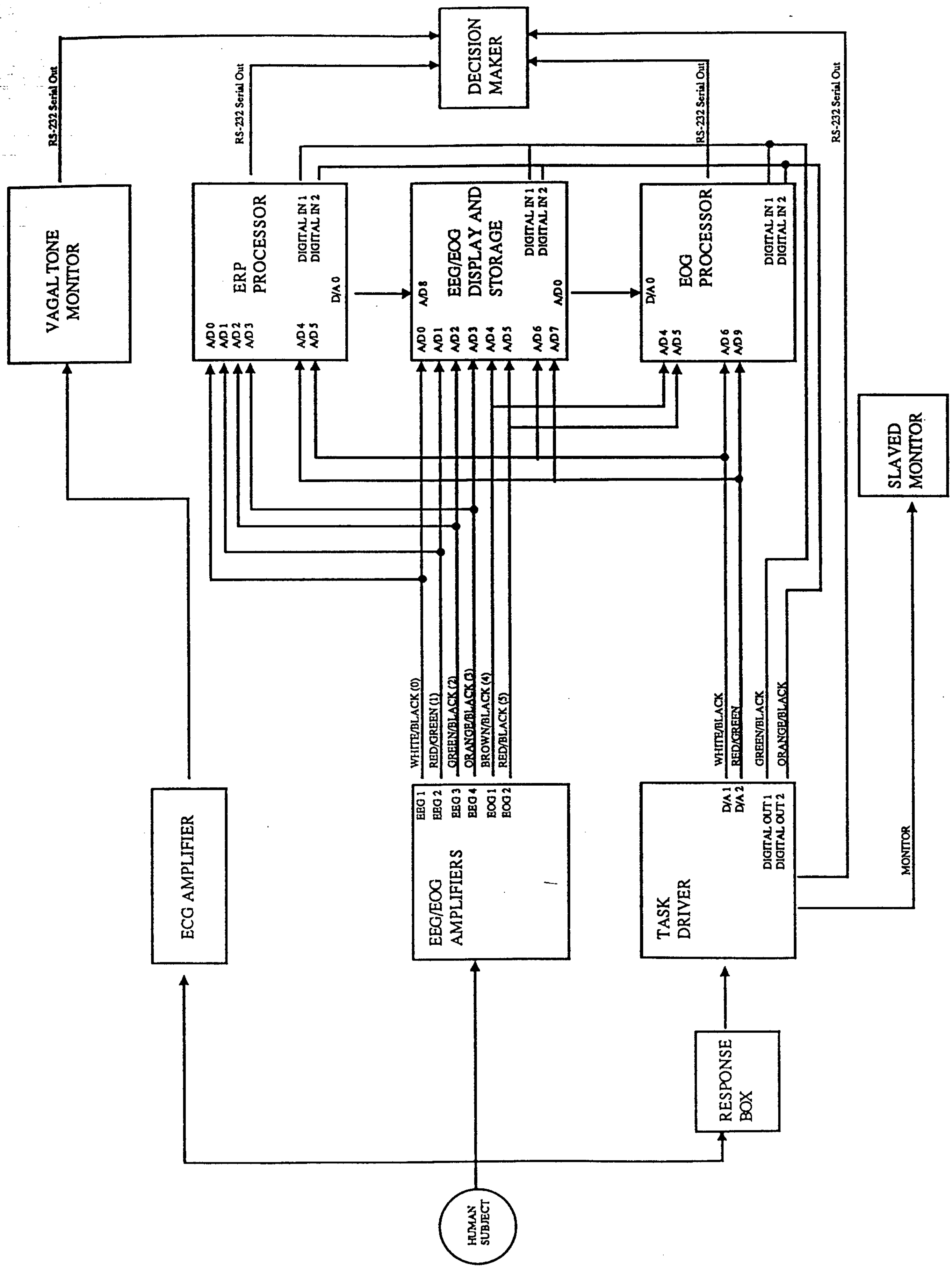



Figure 1 -- A schematic overview of the test-bed system configuration

**Figure 2** -- A schematic of the physical interconnections by which the test-bed was implemented.



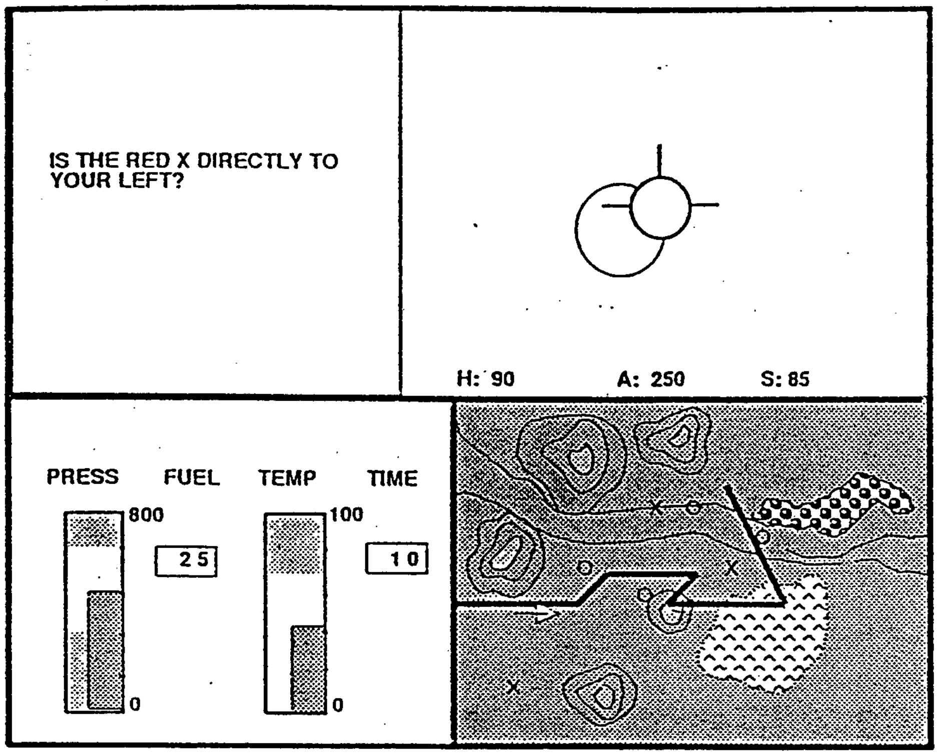
to interactively control which derived measures or trends were displayed at a given time and what decision criteria set-points were in effect. A separate PC served as a scrolling display of the incoming EEG and EOG channels and stored these raw data on disk for retrospective analyses.

## 3.2 Task-Driver System

As a task environment for the present test-bed, a low-fidelity aviation simulation was chosen. This PC-based simulation, called Window/PANES (Workload/PerformANcEe and Simulation) was developed by the Rotorcraft Human Factors Research Branch at NASA Ames Research Center (Ms. Sandra Hart, Manager). Window/PANES provides an environment in which multiple, concurrent discrete and continuous tasks can be presented. Although the displays represent the flight characteristics of a light aircraft, no knowledge of flying is necessary for a subject to learn how to perform the task. The software is nicely designed to support the experimenter in developing scenarios, to log performance and task condition data to disk in real-time as the subject performs the scenario, and to support the experimenter's retrospective analysis of the data.

The display presented on the PC screen has four fixed windows (see Figure 3): (1) a graphic display of commanded and current speed, heading, and altitude presented in a "heading-up" orientation; (2) a north-up map which can depict geographical features, the flight path, the aircraft's position on the flight path, and additional information added for experimental purposes; (3) one, two, three, or four gauges presented in analog or digital form that can be labelled and scaled according to experimenter-defined specifications; and (4) an area in which alphanumeric messages can be displayed. These messages can be used to instruct or inform the subject about the scenario, or they can be used to present stimuli in the context of a secondary (or tertiary) task, in order to provide additional measures of performance. The content of alphanumeric messages, gauges, and objects on the map can be either related to or independent of the primary manual control task. A response box providing for the subject's inputs contains a two-axis joystick (fore/aft: altitude, right/left: heading), a vertical slide potentiometer (fore/aft: speed) and response buttons that can be assigned different meanings depending on the

Figure 3 -- A representative screen showing the Window/PANES task environment (reprinted from the Window/PANES User Manual developed by the Rotorcraft Human Factors Research Branch at NASA Ames.



structure of a particular scenario.

The behavior of all aspects of the display and task that are not under the subject's control are dependent on a script file which specifies the commanded flight path, the significance and dynamics of each gauge, the appearance of alphanumeric information, and the discrete responses anticipated from the subject. Script files are prepared by the experimenter in advance and are not modified by the subject's responses during a flight. Data files which combine the "condition" information in the script file with the subject's performance (speed and accuracy of responding) are logged to disk at regular intervals for retrospective analyses.

We found the Window/PANES task environment to be extremely flexible, readily usable, and well-configured for the present research purposes. Figure 4 illustrates the information flow through the "Task Driver" Processor in the test-bed, on which the Window/PANES software was run. The behavioral performance measures of interest (tracking error, response time and accuracy for discrete stimuli and abnormal gauge changes) were captured as they became available, time-stamped, and then output periodically via a serial port for use as one set of inputs to the Decision-Maker. A digital output signal was sent to the individual data processors in the test-bed in order to provide a common time marker for the start of the scenario. Analog output signals (D-to-A) were sent to the ERP and EOG processors to code the occurrence of a task-relevant event (stimulus) of interest.

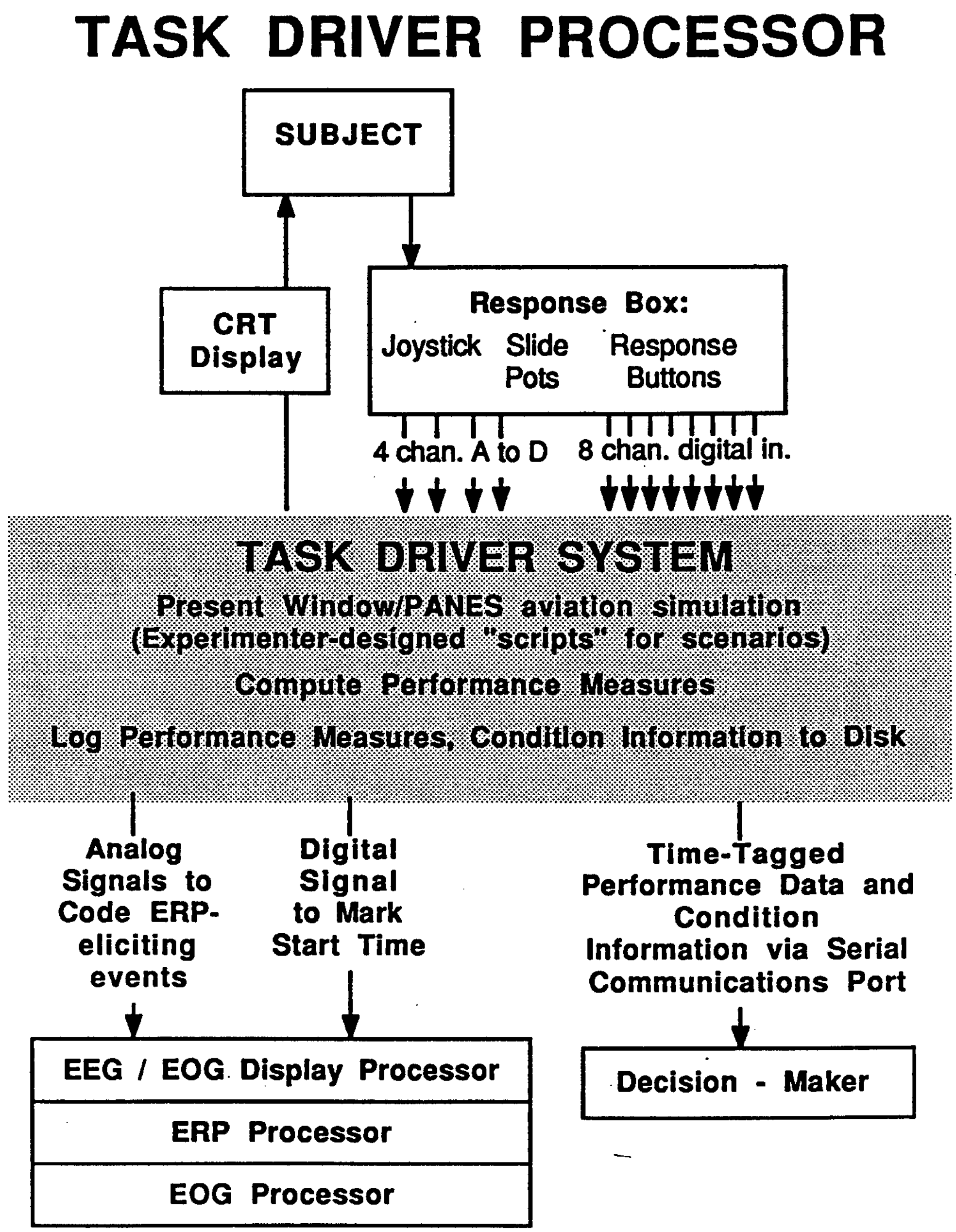
In order to implement Window/PANES for the present project, we made the following changes to the distributed version of the software:

- Adapted the package for use with a Data Translation 2801 interface board instead of the ISAAC board used by the NASA Ames group.

- Altered the display to allow the alphanumeric messages and tracking displays to be partially superimposed, so that subjects could take in both without making saccadic eye movements.

- Added D-to-A output pulses to encode the time of occurrence of specific

Figure 4 -- Information flow through the Task Driver Processor.



events in the scenario that can be used to time-lock ERPs.

- Added a digital output pulse which is issued at the beginning of a scenario to the other processors in the test-bed. This pulse served as a "start" signal, so that the derived measures produced by all processors were time-stamped according to a common reference time.

- Added serial communications routines (Greenleaf CommLiB) to capture the behavioral measures that are streamed to disk by the Window/PANES program in order to send them in near real-time to the Decision Maker.

### 3.3 Justification for the Physiological Measures of Interest

The most informative physiological measures to the investigator interested in mental workload have proven to be heart rate and heart rate variability, eye blink frequency, duration, and latency, and ERP amplitude, latency and scalp distribution for selected components of the waveform. A brief justification for each of these physiological indices is presented below.

Vagal tone and Heart Rate. Heart rate is an often used measure of cardiovascular activation and has been shown by numerous investigators to vary with cognition and mental workload. Vagal tone refers to a derived measure of heart-rate variability that accurately quantifies respiratory sinus arrhythmia. The inter-beat intervals of the heart define a complex time series that reflect a variety of influences on the heart. Respiratory sinus arrhythmia (RSA) is one such influence that has received considerable attention because it is a source of central nervous system mediation of the heart beat. A variety of analytical approaches, including spectral analyses (e.g., Kamphuis & Frowein, 1985; Veldman, et al., 1985), have been applied to heart-rate variability data in an effort to estimate RSA. Unfortunately, spectral analysis makes some untenable assumptions (e.g., that the time series exhibits stationarity) about the statistical properties of the heart beat time series. The vagal tone measure is derived by applying a moving polynomial filter to the beat-to-beat interval data in order to estimate and remove the slowly shifting baseline on which the RSA signal is superimposed (see Porges, 1985). This method dynamically fits a trend line with regression techniques to enable accurate quantification of RSA. Vagal tone appears to be quite sensitive to fluctuations in attention and cognitive load (e.g., Dellinger, et. al., 1986 and unpublished data of our own).

ERP Amplitude, Latency, and Scalp Distribution. Many studies by Donchin and colleagues (see review in Munson, et. al., 1988) have examined the amplitude of the P300 component of the scalp-recorded event-related potential (ERP) as an index of cognitive expectancy and stimulus evaluation. P300 amplitude increases to the extent that processing resources are being devoted to the stimulus of interest, as opposed to stimuli that are occuring in competing tasks. The present investigators have confirmed the sensitivity of ERPs in a series of studies completed for NASA (Horst, et. al. 1984, 1985, 1987). In addition to obtaining P300 results that were consistent with those in the literature, we found two negative-going waveform components that were related to workload. Different components in the waveforms reflected the effects of selective attention, workload, and recognition of a target stimulus (Horst, et. al., 1989).

Scalp-recorded ERPs are usually extracted from the ongoing EEG by signal averaging the brain activity recorded in response to numerous occurrences of an eliciting stimulus. The waveform that is recorded in response to a single such event is referred to here as a "single-trial." The number of trials typically included in time-locked averages range from several dozen to several thousand, depending upon the type of ERP activity of interest and its signal-to-noise ratio with respect to the background EEG. For ERP components related to cognition (e.g., see Donchin, et al., 1978), several dozen trials are typically averaged.

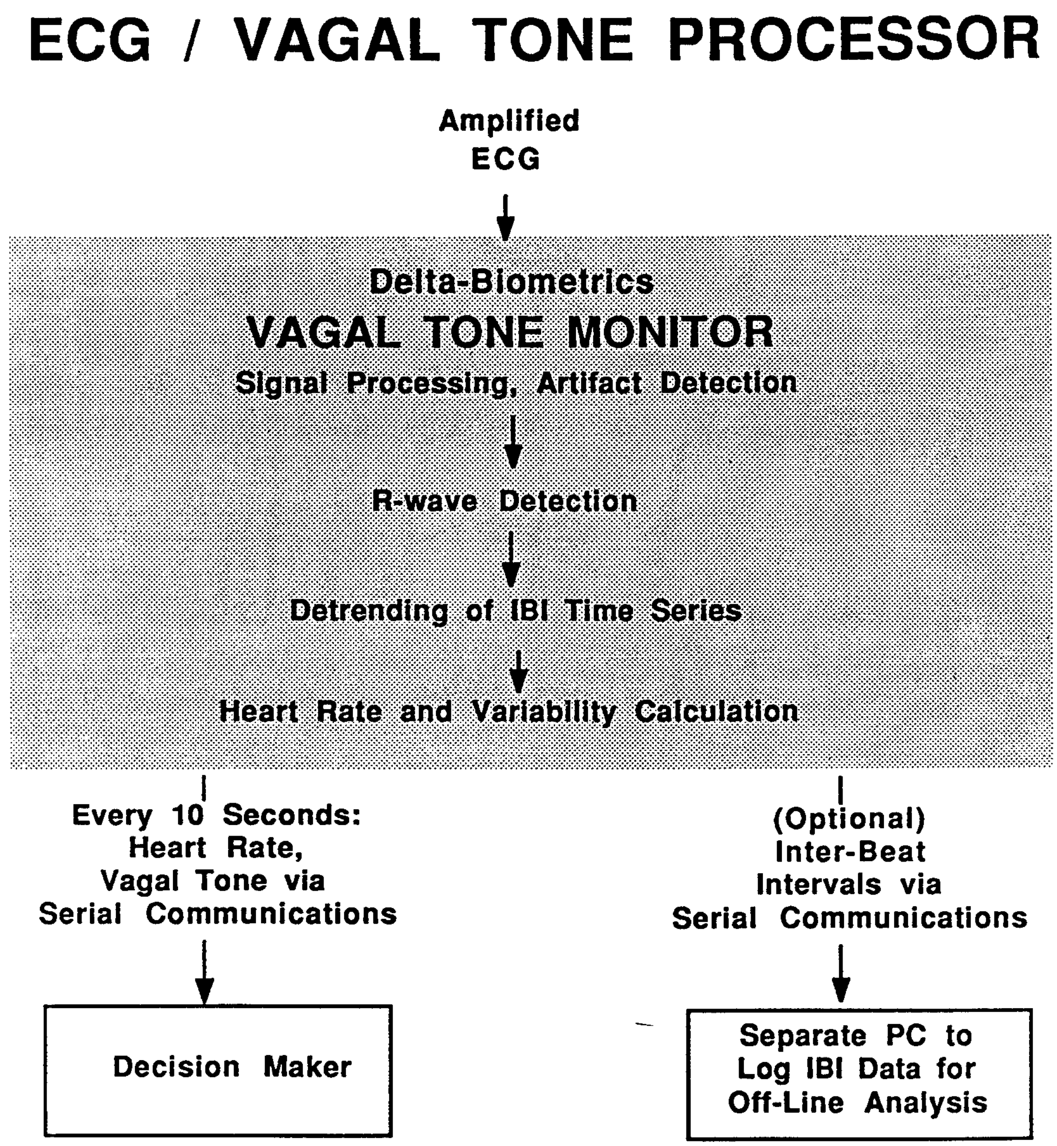
For many applications in operational settings, such signal averaging is impractical, because the eliciting conditions can change unpredictably and one wishes to quantify ERP activity on a moment-to-moment basis. Even if stimulus conditions can be presented repeatedly, as they can for some "open-loop" studies in which system design issues are being addressed and the data can be analyzed off-line, there is a problem with ERP components varying in latency from trial-to-trial. Because the components of interest are related to cognitive processes, and in complex tasks the timing of these processes may vary from trial-to-trial, the ERP components may "jitter" in time from one trial to the next. Average ERPs calculated under such conditions will contain broader, lower amplitude waveshapes than were present on the single-trials that were averaged together. Several techniques have been developed to extract useful information from ERPs on a single-trial basis, including Step-Wise Discriminant Analysis (SWDA), a cross-correlational approach developed by Woody (1967), and a related approach developed by Aunon and McGillem (1975). More recently, Gratton and colleagues (Gratton, et. al., 1989, a,b) have developed a vector filter approach for taking scalp distribution information into account in the single-trial pattern recognition process.

Blink frequency, duration. Stern and colleagues (Bauer, et. al., 1985; Goldstein, et. al., 1985) have demonstrated the usefulness of eye closure frequency and duration as an index of vigilance and mental workload during visual tasks. The present investigators have confirmed the sensitivity of this measure in a recent Army-funded laboratory study of individual differences in workload. We have also implemented a simple pattern recognition algorithm for detecting blinks and for characterizing blink duration. Two criteria are applied at each point in the EOG channel to determine whether a blink occurred at each point in time. First, a blink is characterized by a change in the digitizer values of at least a user-defined magnitude within a user-defined number of time points. This change occurs with a user-specified polarity. Second, a blink is defined by a return of the digitizer values to their original values after a user-defined number of time points within a user-defined tolerance. When both of these criteria are satisfied, a blink is reported at the location of the change detected by the first criterion.

### 3.4 ECG Processor

A Delta-Biometrics Vagal Tone Monitor (VTM) was used as the ECG processor. This is a commercially available system that calculates heart rate and vagal tone from an ECG signal. Vagal tone refers to a derived measure of heart-rate variability that quantifies respiratory sinus arrhythmia and which avoids some of the statistically untenable assumptions of power spectral analyses applied to ECG inter-beat-interval (IBI) data (see Porges, 1985). Figure 5 illustrates the information flow through the ECG Processor in our test-bed.

Figure 5 -- Information flow through the ECG/Vagal Tone Processor.



The VTM was configured to run in its five-second turn-around mode. Thus, it took five seconds of amplified ECG from chest electrodes, screened it for artifact, detected the R-waves in the ECG, calculated IBIs, detrended the IBI time series to filter out frequencies unrelated to respiratory sinus arrhythmia, and calculated heart rate and vagal tone from this corrected time series. These derived measures, along with a time-stamp, were output (approximately once every ten seconds) over a serial interface for use as another set of inputs to the Decision-Maker.

### 3.5 EOG Processor

EOG was amplified by conventional means from sites above and below one eye. Figure 6 illustrates the information flow through the EOG Processor in our test-bed. Five second segments of ongoing EOG were processed in a double buffering scheme. As one five-second period of data was digitized and stored in direct access memory, the preceding five-second's worth of data was processed. The buffer to be processed was digitally filtered and then subjected to an algorithm that applied various experimenter-defined criteria to detect the occurrence of a blink.

Figure 7 illustrates the concept used for pattern recognition of blinks. The algorithm searched for a rising edge that exceeded "change criterion" within some "change time" window. Then after a specified "return time" the algorithm checked to see if the waveform had fallen below the "return criterion" amplitude. If it had, a blink was declared. The peak time of the blink was then found and the half-amplitude on the rising edge was derived. The duration of the blink was determined to be the time from when this half-amplitude point occurred on the rising edge until the waveform had returned to this same amplitude on the falling edge.

Once a blink had been detected, it was time-stamped relative to the digital pulse received from the Task Driver at the beginning of the scenario, its duration was estimated from the rise and fall times that were detected by the pattern recognition routine, and the latency of its peak was determined relative to the last analog event-marker received from the Task Driver. These derived measures were output asynchronously (i.e., only after a blink had been

Figure 6 -- Information flow through the EOG Processor.

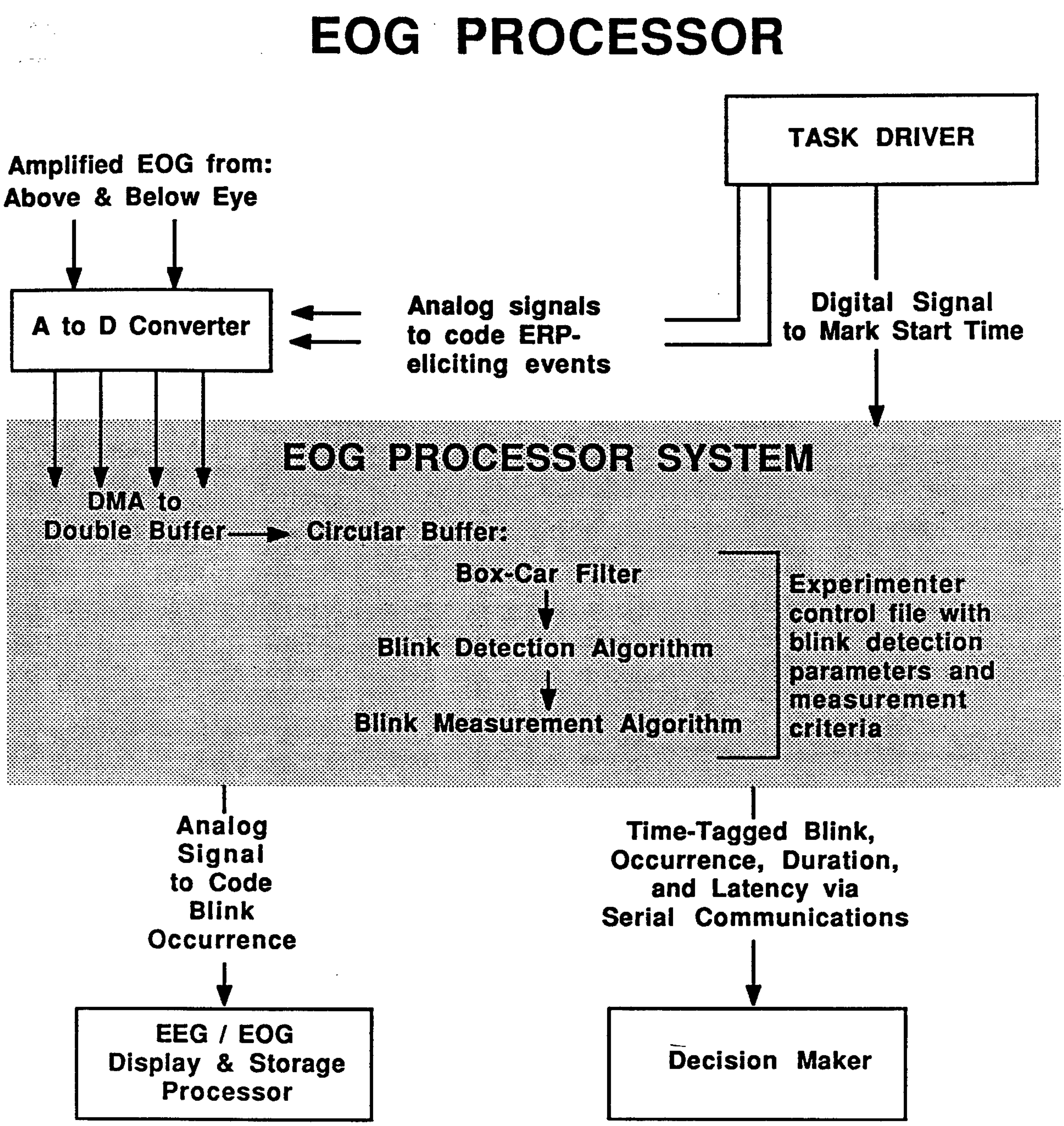
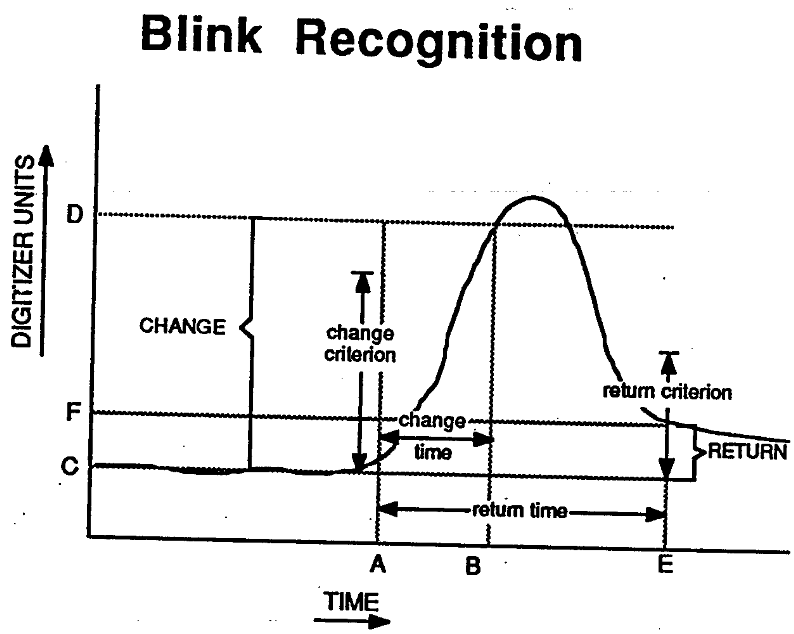


Figure 7 -- A representative EOG trace showing the various criteria by which the algorithm recognized and characterized blinks.



detected) over a serial interface for use as another set of inputs to the Decision-Maker.

### 3.6 ERP Processor

EEG was amplified by conventional means from four scalp sites (Fz, Cz, Pz, and Oz in the International Ten-Twenty System). Of primary interest were several endogenous components of the ERP -- the N200, P300, and Slow Wave, all of which have been shown to reflect cognitive processing (see Donchin, et. al., 1978). Approaches for extracting ERPs from the ongoing EEG had been explored on a previous project (unpublished as yet) and the usefulness of the Vector Filter developed by Gratton, et al. (1989, a, b) was confirmed. For the present feasibility demonstration, it was assumed that the timing of the eliciting stimuli were known to the data processing system. Thus the present ERP components were extracted from ongoing EEG by applying a relatively simple search algorithm to an appropriately filtered segment of multi-channel EEG, time-locked to the events of interest that occurred on the Task Driver.

Figure 8 illustrates the information flow through the ERP Processor in our test-bed. The ERP Processor functioned analogously to the EOG Processor, except that instead of searching the incoming five-second buffers for blinks, it searched the event-marker channel for the occurrence of an event of interest. Having found one, it time-stamped the occurrence relative to the digital signal received from the Task Driver at the beginning of the scenario. The EEG was then digitally filtered and a weighted combination of the data across the various scalp sites was derived for each component of interest, using the Vector Filter. Peaks in these weighted combination waveforms were identified within certain latency epochs, and the peak amplitudes and latencies were calculated. These derived measures were output over the serial interface for use as another set of inputs to the Decision-Maker.

### 3.7 Decision-Maker Processor

Information flow through the Decision-Maker is shown in Figure 9. As mentioned previously, the Decision-Maker received derived performance measures from the Task Driver, ECG Processor, EOG Processor, and ERP Processor, all via serial

Figure 8 -- Information flow through the ERP Processor.

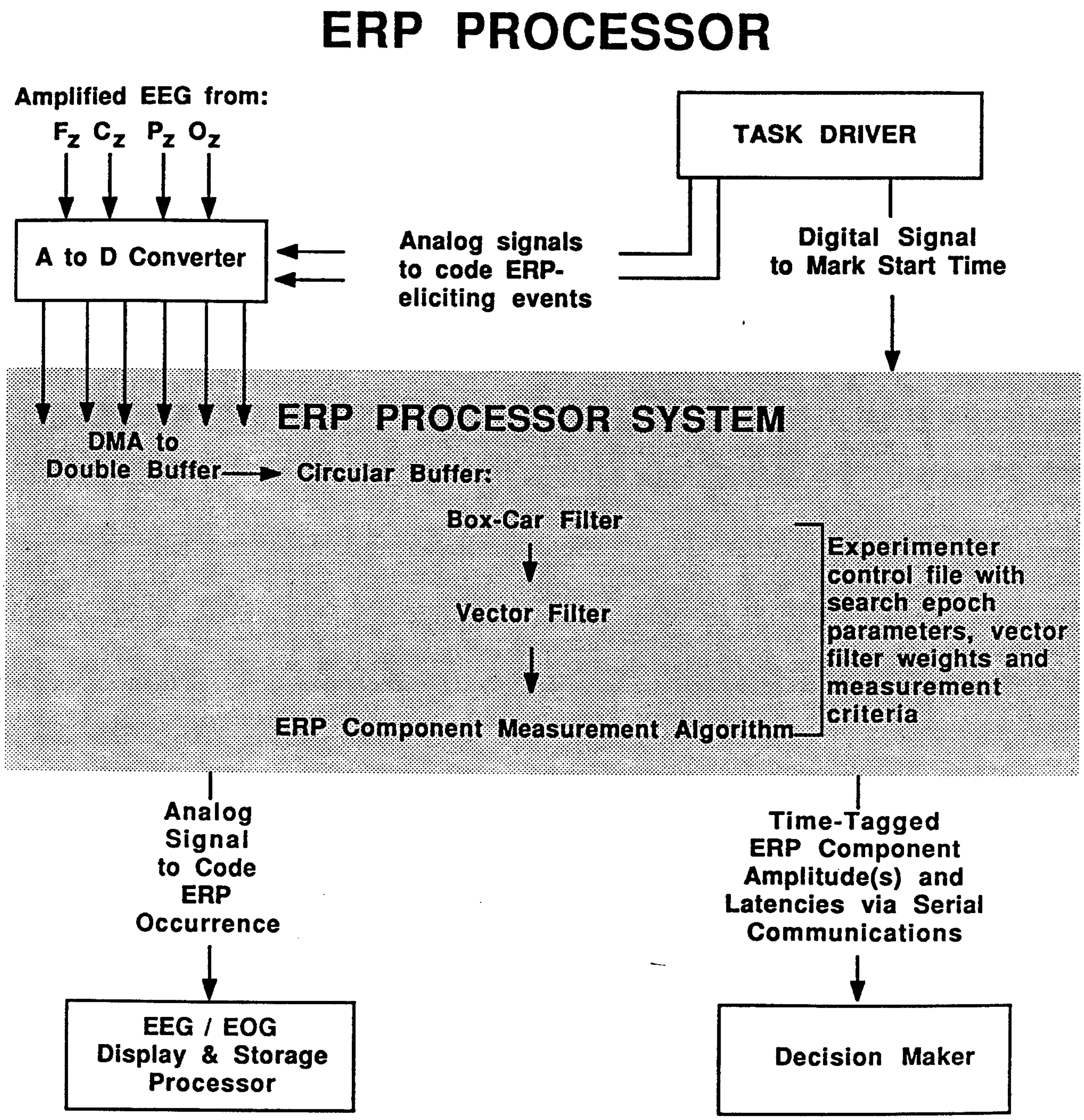
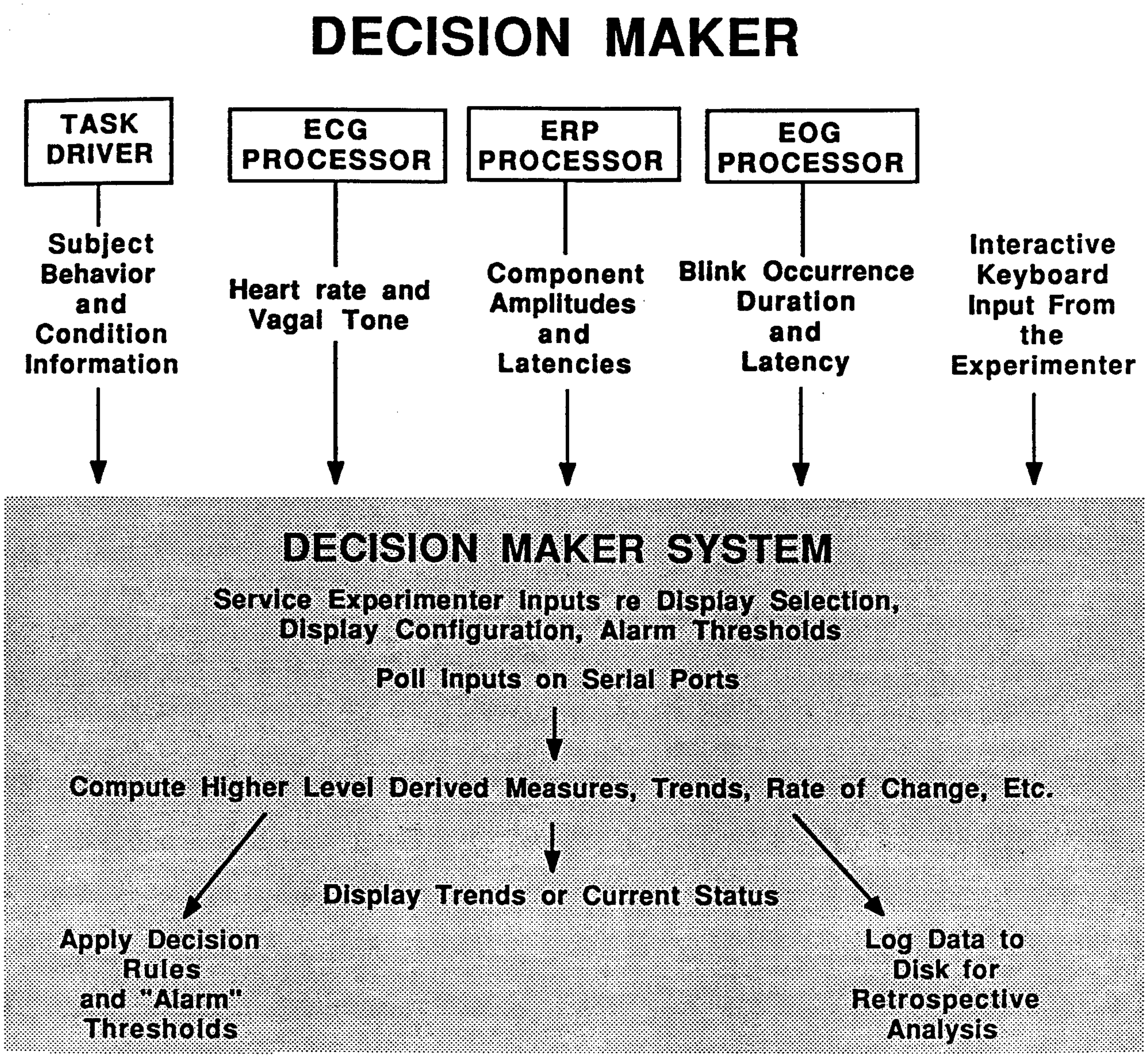


Figure 9 -- Information flow through the Decision-Maker Processor.



communications. The Decision-Maker polled and stored these incoming data and implemented some simple pattern recognition algorithms to allow the triggering of "cautions and warnings" when certain measures exceeded pre-selected set-points. The Decision-Maker also provided an interface for the experimenter to interactively control which derived measures or trends were displayed at a given time and what decision criteria set-points were in effect. There were two primary display modes provided by the Decision-Maker -- a "trends" display on which various combinations of selected measures could be called up, and a "current status" display showing the values of all derived measures at the current "slice" in time.

# 4.0 VALIDATION OF ALGORITHMS AND EXPERIMENTAL MANIPULATIONS

## 4.1 The Approach to Validation

There are several aspects to the validation of the present test-bed. One aspect of validation has to do with the choice of physiological measures and the quality of the present implementation of the real-time analysis algorithms. That is, do the indices chosen lend themselves to real-time analysis and, if so, do the present algorithms perform as intended? Another aspect of validation has to do with the sensitivity of these measures to task manipulations presented in the operational context of interest. Are these measures, viewed in appropriate combinations, sensitive to task difficulty manipulations that affect performance in meaningful situations? Finally, even if the above questions can be answered in the affirmative, there may be an issue regarding the usefulness of the test-bed. Does this test-bed approach lend itself to studying research issues and aviation design problems that could affect the design of next generation aircraft or other complex man-machine systems?

## 4.2 Validation Results To-Date

The present scope of work did not allow a thorough examination of all these issues, and what test results have been obtained can not be presented here in any detail. However, as a proof-of-concept, the present test-bed implementation has proven to be very encouraging. The validation testing that has been conducted to-date can be summarized as follows:

- The Window/PANES task difficulty manipulations were effective in that they were associated with reliable changes in the behavioral measures examined. These behavioral changes were readily apparent when data were averaged over several minute blocks, as in a conventional analysis. Moreover, these behavioral trends were usually apparent in the data transmitted to the Decision-Maker, particularly when the task manipulation was a fairly dramatic one.

- In that the Vagal Tone Monitor was used in its commercially available configuration, which has been thoroughly tested by the manufacturer and used by us on previous projects, there was no need to validate the accuracy of its output. However, an important aspect of the present validation was an examination of the extent to which heart rate and vagal tone measures, based on ten-second segments of ECG data, reflected manipulations of the Window/PANES task. While there is obviously some "noise" to be expected in examining such short segments of ECG, the more dramatic task changes were associated with reliable trends in vagal tone (i.e., an inverse relationship between task difficulty and vagal tone).

- The EOG data processing algorithm was validated by examining its performance with simulated EOG data and by comparing the algorithm's handling of actual EOG with a retrospective visual inspection of stored raw data. The derived EOG measures have proven to be somewhat sensitive to task manipulations, although it is expected that more dramatic changes would be apparent in continuous performance scenarios that put a premium on vigilance and (the warding off of) visual fatigue.

- The ERP data processing algorithm was likewise validated with simulated EEG/ERP data and with previously recorded data from a standard "Oddball" task (e.g., see 8). The ability to extract single-trial estimates of ERP amplitude and latency which mirror the differences seen in conventional average ERPs has been impressive. However, the sensitivity of these waveforms, whether averaged or single trial, to subtle task manipulations in the Window/PANES environment has not been impressive to-date. This may, of course, be due to the task scenarios used thus far.

- The present implementation of the Decision-Maker has proven to be very useful for both monitoring the course of data collection in real-time and examining relationships among the various derived measures retrospectively. Much more remains to be done in the way of optimizing the decision rules based on the available derived measures.

## 5.0 POTENTIAL APPLICATIONS AND THEIR FEASIBILITY

In most respects, the performance of the present test-bed has been very encouraging. The validation results obtained to-date suggest that we now have a reasonable ability to detect changes in the derived measures of interest "on the fly." The test-bed ran successfully in near "real-time" -- i.e., with lags of approximately five seconds in the EOG and ERP measures, a lag of approximately ten seconds in the ECG measures, and a lag of approximately one second in the behavioral measures. No serious attempt has yet been made to optimize the buffer lengths that underlie these lags, and it is expected that significant reductions will be possible in these turn-around times. Of course, the approximation to real-time that is achievable will be highly dependent on the speed of the processors on which the algorithms are implemented. The usefulness of the present combination of measures was suggested by an encouraging ability to infer changes in task difficulty of the Window/PANES task. This inference ability has thus far only been examined with parameters for the various pattern recognition algorithms being tailored to the individual subject (not an unreasonable constraint in systems that will be operated by a relatively small group of highly trained specialists). Furthermore, the initial testing conducted to-date has concentrated only on rather dramatic manipulations in task difficulty.

Much more needs to be done in examining the performance of this test-bed and in optimizing the parameters used in the pattern recognition algorithms and decision rules. However, the present implementation has confirmed the feasibility of the approach and suggested the usefulness of the test-bed for future research on dynamic man-machine interactions. Potential future directions for the use of this test bed include the following:

- Further validation of the usefulness of the present behavioral and physiological measures. These efforts should include additional systematic collection of empirical data, as well as more fine-grained manipulations in task difficulty and subject status (e.g., sleep deprivation, drug effects, continuous performance challenges).

- Using the test-bed to develop effective decision rules for inferring operator mental status. This effort should explore inference rules based on contingencies among the measures and the generalizability of these rules within and between test subjects.

- Integration of test-bed functions into a single, multi-tasking workstation. This integration effort should include attempts to optimize the real-time algorithms and an exploration of alternative designs for implementing distributed processing approaches that are analogous to those achieved with the present multi-processor configuration.

- Finally, a number of uses are foreseen for the present test-bed in facilitating future research and development on the role of the human operator in automated and semi-automated systems. The test-bed should lend itself to studies of "closed-loop" decision-aiding, dynamic man-machine task allocation, and computational approaches to recognizing operationally significant patterns, across time, in multiple performance measures.

## 6.0 REFERENCES


Aunon, J. I. & McGillem, C. D. (1975). Techniques for processing single evoked potentials. Transactions of the San Diego Biomedical Symposium, 14, 211-218.

Bauer, L. O., Goldstein, R., & Stern, J. A. (1987). Effects of information processing demands on physiological response patterns. Psychophysiology, 29, 213-234.

Bauer, L. O., Strock, B. D., Goldstein, R., Stern, J. A., & Walrath, L. C. (1985). Auditory discrimination and the eyeblink. Psychophysiology, 22, 636-641.

Dellinger, J. A., Taylor, H. L., & Porges, S. W. (1986). Atropine sulfate effects on aviator performance and on respiratory-heart period interactions. Aviation, Space and Environmental Medicine.

Donchin, E., Kramer, A, & Wickens, C. (1986). Applications of brain event-related potentials to problems in engineering psychology. In M. G. H. Coles, E. Donchin, & S. W. Porges, (Eds.), Psychophysiology: Systems, Processes, and Applications, New York: Guilford, 702-718.

Donchin, E., Ritter, W. & McCallum, W. C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), Event-related Brain Potentials in Man, New York: Academic Press, 349-412.

Goldstein, R., Walrath, L. C., Stern, J. A., & Strock, B. D. (1985). Blink activity in a discrimination task as a function of stimulus modality and schedule presentation. Psychophysiology, 22, 629-635.

Gopher, D., & Donchin, E. (1986). Workload -- An examination of the concept. In K. Boff & L. Kaufman (Eds.), Handbook of perception and human performance. New York: Wiley & Sons.

Gratton, G., Coles, M. G. H., & Donchin, E. (1989). A procedure for using multi-electrode information in the analysis of components of the event-related potential: Vector filter. Psychophysiology, 26, 222-232.

Gratton, G., Kramer, A. F., Coles, M. G. H., & Donchin, E. (1989). Simulation studies of latency measures of components of the event-related potential. Psychophysiology, 26, 233-248.

Hart, S. G. (1986). Theory and measurement of human workload. In J. Zeidner (Ed.), Human productivity enhancement: Cognitive processes in system design. New York: Praeger Scientific.

Horst, R. L. (1988). An overview of current approaches and future challenges in physiological monitoring. In J. R. Comstock, Jr. (Ed.), Mental-State Estimation -- 1987, NASA Conference Publication 2504, 25-42.

Horst, R. L., Mahaffey, D. L., Munson, R. C. (1989). Brain-wave Measures of

Workload in Advanced Cockpits: The Transfer of Technology from Laboratory to Cockpit Simulator, NASA Contractors Report 4240.

Horst, R. L., Munson, R. C., & Ruchkin, D. S. (1984). Event-related potential indices of workload in a single task paradigm. Proceedings of the Human Factors Society -- 28th Annual Meeting, San Antonio, Texas, 727-731.

Horst, R. L., Munson, R. C., Ruchkin, D. S. (1985). Processing inferences derived from event-related potential measures in a monitoring task. Proceedings of the Human Factors Society -- 29th Annual Meeting, Baltimore, Md., 976-980.

Horst, R. L., Ruchkin, D. S., & Munson, R. C. (1987). ERP processing negativities related to workload. In R. Johnson, Jr., R. Parasuraman, & J. W. Rohrbaugh (Eds.), Current Trends in Event-Related Potential Research (EEG Supplement 40), Amsterdam: Elsevier Science Publishers, 186-190.

Johnson, L. C. (1980). Use of physiological measures to monitor operator state. In F. W. Gomer (Ed.), Biocybernetic applications for military systems (Report No. MDC E2191). St. Louis: McDonnell Douglas Astronautics Company.

Kamphuis, A., & Frowein, H. W. (1985). Assessment of mental effort by means of heart rate spectral analysis. In J. F. Orlebeke, G. Mulder, & L. J. P. van Doornen (Eds.), Psychophysiology of cardiovascular control. New York: Plenum Press, 841-854.

Moray, N. (Ed.). (1979). Mental workload: Its theory and measurement. New York: Plenum Press.

Munson, R. C., Horst, R. L. & Mahaffey, D. L. (1988). Primary task event-related potentials related to different aspects of information processing. In J. R. Comstock, Jr. (Ed.), Mental-State Estimation 1987, NASA Conference Publication 2504, 163-178.

Orlebeke, J. F., Mulder, G. & van Doornen, L. J. P. (Eds.) (1985). Psychophysiology of Cardiovascular Control. New York: Plenum Press.

Porges, S. W. (1985). Respiratory sinus arrhythmia: An index of vagal tone. In J. F. Orlebeke, G. Mulder, & L. J. P. van Doornen (Eds.), Psychophysiology of Cardiovascular Control, New York: Plenum Press, 437-450.

Veldman, J. B. P., Mulder, L. J. M., Mulder, G., & van der Heide, D. (1985). Attention, effort and sinus arrhythmia: How far are we? In J. F. Orlebeke, G.

Woody, C. D. (1967). Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Medical & Biological Engineering, 5, 539-553.

Zacharias, G. L. (1980). Physiological correlates of mental workload (Report No. 4308). Bolt Beranek and Newman, Inc.

# APPENDIX A

## DESCRIPTION OF THE WINDOW/PANES TASK ENVIRONMENT

(Note that this description is adapted from the Window/PANES User Manual, developed and written by the Rotorcraft Human Factors Branch at NASA Ames Research Center; Ms. Sandra Hart, Manager)

Window/PANES (Workload/PerformANcE Simulation) is a simple simulation of a flying task designed as a tool with which to conduct research on the effects of complex task structure and subtask demands on workload, training, and performance. It was designed and developed by the Rotorcraft Human Factors Branch at NASA Ames. Window/PANES provides an environment in which multiple, concurrent discrete and continuous tasks can be presented. The content of alphanumeric messages, gauges, and objects on a map can be either related to or independent of the primary manual control task. Although the displays represent the flight characteristics of a light aircraft, no knowledge of flying should be necessary for a subject to learn how to perform the task.

Window/PANES runs on an IBM/AT (or compatible) with a high resolution graphics display. A Cyborg ISAAC interface (or equivalent) is used for A/D conversion and timing. The display has four fixed windows: (1) a graphic display of commanded and current speed, heading, and altitude presented in a "heading-up" orientation; (2) a north-up map which can depict geographical features, the flight path, the aircraft's position on the flight path, and additional information added for experimental purposes; (3) 1, 2, 3, or 4 gauges presented in analog or digital form that can be labelled and scaled according to the whims of the experimenter; and (4) an area where alphanumeric messages can be displayed.

A response box is used by the subjects in performing a Window/PANES scenario. It contains a two-axis joystick (fore/aft: altitude, right/left: heading) and vertical slide pot (fore/aft: speed) used for vehicle control. Six response buttons are provided which can be assigned different meanings depending on the structure of a particular scenario.

The behavior of everything that is not under the subjects' control (e.g.

movement of the target on the map, alphanumeric messages, and changes in the values of each gauge) is controlled by a script file which specifies the commanded flight path, the significance and dynamics of each gauge, the appearance of alphanumeric information, and the discrete responses anticipated from the subject. Script files are prepared by the experimenter in advance and are not modified by subjects' responses during a flight.

The subject's primary task is to minimize the distance between the position of his own aircraft and the target (e.g., heading and altitude deviation) and the difference between the size of his own aircraft and the target (speed deviation). The "low frequency" tracking task involves following changes in heading, speed, and altitude programmed into the target's flight path. The number and frequency of flightpath changes can be manipulated to create scenarios (or segments within a scenario) with different levels of difficulty. If no changes in altitude, heading, or speed are programmed, the flight path control task closely resembles a traditional two-axis laboratory tracking task. A "high frequency" tracking task (e.g., quasi random "noise") can be superimposed on either heading and/or altitude. High frequency disturbances may not be applied to speed. Thus, depending on the combination of low- and high-frequency tracking task conditions selected, a one- or two-axis tracking task can be created with either low or high response demands on either axis. In addition, subjects may also be required to acknowledge or evaluate information presented on the map, values shown on the gauges, or alphanumeric messages by pressing one of the buttons. Again, the frequency with which responses are required and the difficulty of the discrete tasks is determined by the experimenter.

## EXPERIMENTAL DISPLAY

The screen is divided into four parts: (1) The upper left window displays alphanumeric information; (2) the lower left window display up to 4 gauges; (3) the upper right window displays the tracking task; and (4) the lower right window displays the map.

### Alphanumeric Window

Alphanumeric characters are displayed in a 3-line x 25-character area in the upper left window. Any text string entered by the experimenter can be displayed for as long as the experimenter specifies in the script that controls each experimental scenario. The following examples demonstrate how this window might be used to display different types of experimental tasks:

Gauges

There may be 1, 2, 3, or 4 gauges displayed in analog or digital form in the lower left window. A 4-character title may be displayed at the top of each gauge. Analog gauges are presented as a white bar, (115 pixels, 2.1 in) with a moving blue indicator controlled by commands in the script. Digital gauges are three-digit numbers that "cycle" as their values are changed by commands in the script. The subject may be required to monitor one or more gauges and respond when the pointer moves into an area designated as a "red" or danger zone (this may be displayed explicitly, or not), when a numeric value that is out of bounds appears, or compute new values based on the information provided by the gauges. The meaning attached to the gauges and the values they assume may be related to the flight task or it may be completely independent.

Tracking Window

The tracking window represents a pilot's view out the window of an aircraft; a "heading-up" display of clear blue sky. Ownship is always displayed in the center (analogous to the aircraft symbol painted in the center of an attitude indicator). The "flight plan" or commanded flight path is depicted by a circular symbol that varies in position and size.

Ownship is represented by a symbol that is fixed at 11 pixels in diameter. This represents 2.5 deg (horizontally) and 12.42 ft (vertically), given the scale selected for the tracking display. The subject's task is to superimpose the two symbols, so that there is no difference in either position (heading and altitude) or size (speed). The target's size ranges from 2 to 20 pixels in diameter in 1-pixel (3-kt) increments. A decrease in target size relative to the size of the ownship symbol indicates that the target is moving away from (or going faster) than ownship. An increase in target size relative to the size of the ownship symbol indicates that the commanded speed is less than the ownship speed. The maximum range of displayed speed differences ranges from +27 knots to -27 knots. Target size does not change beyond the maximum values (e.g., 2 or 20 pixels) for greater deviations.

### Map Display

The map area depicts a scene created by the experimenter. It may represent realistic geographical features (such as mountains, rivers, lakes, forests), and include additional symbols, alphanumeric characters, or waypoints to increase the face validity and complexity of the task. Or, it may depict a grid, stylized objects that vary in size, shape, or color, or a uniform texture. Physically, the map area is 4.87 in wide (174 pixels) and 3.5 in high (91 pixels). It is scaled to represent a geographical area that is 40 by 25 mi. The target flight path may be superimposed on the map or not, as the experimenter chooses. The maps are created with Dr. Halo graphics and stored prior to an experiment. They may be given any name and need have no special extension. Several different maps may be used with the same script to create identical experimental conditions that only appear to be different or the same map may be used with a number of different scripts.

## EXPERIMENTAL SCRIPTS

For each scenario,the experimenter must prepare a script file that contains all of the necessary information: a header that contains initialization and setup information and "time line: that contains a detailed description of the events that will transpire.

"Noise files" which contain 8000 filtered random numbers are generated off line and used to impose disturbances on the heading and altitude control tasks, if the experimenter chooses. Noise files are named by the experimenter and need no special extension, although one can be added for clarity (e.g., noise.alt, noise.hdg). The difficulty of the control task is determined by (1) the number of changes in speed, heading or altitude and (2) the bandwidth and amplitude of the"noise file" superimpose on the heading and/or altitude. If no changes are programmed, and no disturbances are added to one axis, a two-axis tracking task is created. If no changes or disturbances are programmed for any axis, the tracking portion of the program is effectively inhibited. This option can be useful in obtaining single-task baseline measures for different discrete tasks.

"Waypoints" are used to segment the discrete and continuous control performance measures for data analysis. Waypoints usually represents a point in the script when a change in speed, heading, or altitude are specified. Waypoints may also bracket experimentally significant segments of flight even if no change in speed, heading or altitude are desired. In this case a dummy "waypoint" is created. A "new" value for one flight parameter is entered that is the same as the current value (speed or altitude are influenced most directly, so they would be preferable). This will not change the current value, but will effectively initiate a new segment for data analysis. A second dummy "waypoint" can be entered to end the segment.

The event data associated with a particular scenario and the performance of a particular subject are stored in the named event data file. Event data are time-tagged and provide a record of which response buttons the subject should have pressed, the buttons which were pressed, and the time at which gauges reach values which were specified in the script. These data may be used to perform reaction time and percentage correct data analyses.

## DATA OUTPUT FILES

There are three different data output files associated with each scenario flown. The <u>event data</u> output file combines the information in the script file with the subject's performance. Every significant event and its associated time is written to the vent file. The <u>waypoint data</u> file records the deviations of the target's position, for heading, altitude and speed, relative to the ownship, every 100 msecs. The <u>scenario data</u> file records the flight scenario as it was specified in the script file. Programs have been developed to analyze the data which have been stored in these data files.

### Inter-Waypoint Statistics

Inter-waypoint statistics (IWS) allows the experimenter to obtain data on tracking averaged for specific intervals during a scenario. The IWS program uses the waypoint data generated during the flight to produce the tracking data averaged over these intervals.

View

VIEW allows the experimenter to view (and show to the subject) the deviations from the target that his aircraft flew during the course of the flight. The information is presented in a graphic format with the three controlled axes presented in three graphs on one screen, stacked one above the other. The center line of each chart represents ideal tracking performance (no error) and the deviations about this line indicate the actual performance of the subjects aircraft throughout the course of the flight.

As described in the text, ARD Corporation made several modifications to the Window/PANES software for the present study. These modifications included adapting the I/O interface for use with a Data Translation 2801 board instead of the ISAAC interface, incorporating Greenleaf Software Inc. serial communications routines in order to send behavioral data and condition information to the Decision-Maker, and modifying the tracking task display to present color changes in the "ownship" symbol as stimuli for a choice reaction time task.

## APPENDIX A

## DESCRIPTION OF THE WINDOW/PANES TASK ENVIRONMENT

(Note that this description is adapted from the Window/PANES User Manual, developed and written by the Rotorcraft Human Factors Branch at NASA Ames Research Center; Ms. Sandra Hart, Manager)

Window/PANES (Workload/PerformANcE Simulation) is a simple simulation of a flying task designed as a tool with which to conduct research on the effects of complex task structure and subtask demands on workload, training, and performance. It was designed and developed by the Rotorcraft Human Factors Branch at NASA Ames. Window/PANES provides an environment in which multiple, concurrent discrete and continuous tasks can be presented. The content of alphanumeric messages, gauges, and objects on a map can be either related to or independent of the primary manual control task. Although the displays represent the flight characteristics of a light aircraft, no knowledge of flying should be necessary for a subject to learn how to perform the task.

Window/PANES runs on an IBM/AT (or compatible) with a high resolution graphics display. A Cyborg ISAAC interface (or equivalent) is used for A/D conversion and timing. The display has four fixed windows: (1) a graphic display of commanded and current speed, heading, and altitude presented in a "heading-up" orientation; (2) a north-up map which can depict geographical features, the flight path, the aircraft's position on the flight path, and additional information added for experimental purposes; (3) 1, 2, 3, or 4 gauges presented in analog or digital form that can be labelled and scaled according to the whims of the experimenter; and (4) an area where alphanumeric messages can be displayed.

A response box is used by the subjects in performing a Window/PANES scenario. It contains a two-axis joystick (fore/aft: altitude, right/left: heading) and vertical slide pot (fore/aft: speed) used for vehicle control. Six response buttons are provided which can be assigned different meanings depending on the structure of a particular scenario.

The behavior of everything that is not under the subjects' control (e.g.

movement of the target on the map, alphanumeric messages, and changes in the values of each gauge) is controlled by a script file which specifies the commanded flight path, the significance and dynamics of each gauge, the appearance of alphanumeric information, and the discrete responses anticipated from the subject. Script files are prepared by the experimenter in advance and are not modified by subjects' responses during a flight.

The subject's primary task is to minimize the distance between the position of his own aircraft and the target (e.g., heading and altitude deviation) and the difference between the size of his own aircraft and the target (speed deviation). The "low frequency" tracking task involves following changes in heading, speed, and altitude programmed into the target's flight path. The number and frequency of flightpath changes can be manipulated to create scenarios (or segments within a scenario) with different levels of difficulty. If no changes in altitude, heading, or speed are programmed, the flight path control task closely resembles a traditional two-axis laboratory tracking task. A "high frequency" tracking task (e.g., quasi random "noise") can be superimposed on either heading and/or altitude. High frequency disturbances may not be applied to speed. Thus, depending on the combination of low- and high-frequency tracking task conditions selected, a one- or two-axis tracking task can be created with either low or high response demands on either axis. In addition, subjects may also be required to acknowledge or evaluate information presented on the map, values shown on the gauges, or alphanumeric messages by pressing one of the buttons. Again, the frequency with which responses are required and the difficulty of the discrete tasks is determined by the experimenter.

## EXPERIMENTAL DISPLAY

The screen is divided into four parts: (1) The upper left window displays alphanumeric information; (2) the lower left window display up to 4 gauges; (3) the upper right window displays the tracking task; and (4) the lower right window displays the map.

### Alphanumeric Window

Alphanumeric characters are displayed in a 3-line x 25-character area in the upper left window. Any text string entered by the experimenter can be displayed for as long as the experimenter specifies in the script that controls each experimental scenario. The following examples demonstrate how this window might be used to display different types of experimental tasks:

Gauges

There may be 1, 2, 3, or 4 gauges displayed in analog or digital form in the lower left window. A 4-character title may be displayed at the top of each gauge. Analog gauges are presented as a white bar, (115 pixels, 2.1 in) with a moving blue indicator controlled by commands in the script. Digital gauges are three-digit numbers that "cycle" as their values are changed by commands in the script. The subject may be required to monitor one or more gauges and respond when the pointer moves into an area designated as a "red" or danger zone (this may be displayed explicitly, or not), when a numeric value that is out of bounds appears, or compute new values based on the information provided by the gauges. The meaning attached to the gauges and the values they assume may be related to the flight task or it may be completely independent.

Tracking Window

The tracking window represents a pilot's view out the window of an aircraft; a "heading-up" display of clear blue sky. Ownship is always displayed in the center (analogous to the aircraft symbol painted in the center of an attitude indicator). The "flight plan" or commanded flight path is depicted by a circular symbol that varies in position and size.

Ownship is represented by a symbol that is fixed at 11 pixels in diameter. This represents 2.5 deg (horizontally) and 12.42 ft (vertically), given the scale selected for the tracking display. The subject's task is to superimpose the two symbols, so that there is no difference in either position (heading and altitude) or size (speed). The target's size ranges from 2 to 20 pixels in diameter in 1-pixel (3-kt) increments. A decrease in target size relative to the size of the ownship symbol indicates that the target is moving away from (or going faster) than ownship. An increase in target size relative to the size of the ownship symbol indicates that the commanded speed is less than the ownship speed. The maximum range of displayed speed differences ranges from +27 knots to -27 knots. Target size does not change beyond the maximum values (e.g., 2 or 20 pixels) for greater deviations.

### Map Display

The map area depicts a scene created by the experimenter. It may represent realistic geographical features (such as mountains, rivers, lakes, forests), and include additional symbols, alphanumeric characters, or waypoints to increase the face validity and complexity of the task. Or, it may depict a grid, stylized objects that vary in size, shape, or color, or a uniform texture. Physically, the map area is 4.87 in wide (174 pixels) and 3.5 in high (91 pixels). It is scaled to represent a geographical area that is 40 by 25 mi. The target flight path may be superimposed on the map or not, as the experimenter chooses. The maps are created with Dr. Halo graphics and stored prior to an experiment. They may be given any name and need have no special extension. Several different maps may be used with the same script to create identical experimental conditions that only appear to be different or the same map may be used with a number of different scripts.

## EXPERIMENTAL SCRIPTS

For each scenario,the experimenter must prepare a script file that contains all of the necessary information: a header that contains initialization and setup information and "time line: that contains a detailed description of the events that will transpire.

"Noise files" which contain 8000 filtered random numbers are generated off line and used to impose disturbances on the heading and altitude control tasks, if the experimenter chooses. Noise files are named by the experimenter and need no special extension, although one can be added for clarity (e.g., noise.alt, noise.hdg). The difficulty of the control task is determined by (1) the number of changes in speed, heading or altitude and (2) the bandwidth and amplitude of the"noise file" superimpose on the heading and/or altitude. If no changes are programmed, and no disturbances are added to one axis, a two-axis tracking task is created. If no changes or disturbances are programmed for any axis, the tracking portion of the program is effectively inhibited. This option can be useful in obtaining single-task baseline measures for different discrete tasks.

"Waypoints" are used to segment the discrete and continuous control performance measures for data analysis. Waypoints usually represents a point in the script when a change in speed, heading, or altitude are specified. Waypoints may also bracket experimentally significant segments of flight even if no change in speed, heading or altitude are desired. In this case a dummy "waypoint" is created. A "new" value for one flight parameter is entered that is the same as the current value (speed or altitude are influenced most directly, so they would be preferable). This will not change the current value, but will effectively initiate a new segment for data analysis. A second dummy "waypoint" can be entered to end the segment.

The event data associated with a particular scenario and the performance of a particular subject are stored in the named event data file. Event data are time-tagged and provide a record of which response buttons the subject should have pressed, the buttons which were pressed, and the time at which gauges reach values which were specified in the script. These data may be used to perform reaction time and percentage correct data analyses.

## DATA OUTPUT FILES

There are three different data output files associated with each scenario flown. The <u>event data</u> output file combines the information in the script file with the subject's performance. Every significant event and its associated time is written to the vent file. The <u>waypoint data</u> file records the deviations of the target's position, for heading, altitude and speed, relative to the ownship, every 100 msecs. The <u>scenario data</u> file records the flight scenario as it was specified in the script file. Programs have been developed to analyze the data which have been stored in these data files.

### Inter-Waypoint Statistics

Inter-waypoint statistics (IWS) allows the experimenter to obtain data on tracking averaged for specific intervals during a scenario. The IWS program uses the waypoint data generated during the flight to produce the tracking data averaged over these intervals.

View

VIEW allows the experimenter to view (and show to the subject) the deviations from the target that his aircraft flew during the course of the flight. The information is presented in a graphic format with the three controlled axes presented in three graphs on one screen, stacked one above the other. The center line of each chart represents ideal tracking performance (no error) and the deviations about this line indicate the actual performance of the subjects aircraft throughout the course of the flight.

As described in the text, ARD Corporation made several modifications to the Window/PANES software for the present study. These modifications included adapting the I/O interface for use with a Data Translation 2801 board instead of the ISAAC interface, incorporating Greenleaf Software Inc. serial communications routines in order to send behavioral data and condition information to the Decision-Maker, and modifying the tracking task display to present color changes in the "ownship" symbol as stimuli for a choice reaction time task.