AGARD-AG

AGARDograph No. 246

Survey of Methods to Assess Workload

NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY
ON BACK COVER

(1)

(14) AGARD-AG-246

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

(12) 62 (11) Aug 79

AGARDograph No. 246

(6) SURVEY OF METHODS TO ASSESS WORKLOAD

Edited by Richard E. McKenzie

(10) Bryce O. Hartman, [redacted]
~~Crew Technology Division~~
USAF School of Aerospace Medicine
Brooks Air Force Base, Texas, 78235
USA

and

Richard E. McKenzie, Ph.D.
San Antonio, Texas, USA

DDC
DEC 14 1979
E

This document has been approved for public release and sale; its distribution is unlimited.

This AGARDograph was sponsored by the AGARD Aerospace Medical Panel.

THE MISSION OF AGARD

The mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Exchanging of scientific and technical information;
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Providing scientific and technical advice and assistance to the North Atlantic Military Committee in the field of aerospace research and development;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced
directly from material supplied by AGARD or the authors.

Published August 1979

ISBN 92-835-1332-0

Printed by Technical Editing and Reproduction Ltd
Harford House, 7–9 Charlotte St, London, W1P 1HD

PREFACE

This multi-authored AGARDograph represents the preliminary survey of the Working Group (AMP-WG-08). "Evaluation of Methods to Assess Workload" was initiated by the AGARD Aerospace Medical Panel in January 1977, following approval by the National Delegates Board (NDB) in the fall of 1976. Working Group meetings were held at Cologne (April 1977), London (October 1977), Fort Rucker, Alabama (May 1978), and Paris (November 1978), concurrent with symposia conducted by the Aerospace Medical Panel. Early meetings focused, as would be expected, on the scope of the task. While it was evident that the broad outline could be described with a high degree of agreement, it was also apparent that tasking individual members with sub-areas would require that they prepare manuscripts de novo, a level of effort clearly not desired, given the substantial burden each of them already had in his own laboratory. It was therefore decided at Fort Rucker to seek contributed chapters from Working Group members and others in the NATO scientific community who had on hand materials which could be readily adapted to the objectives of the Working Group. As the reader will see, numerous contributions were received. The editors feel that the result is a wide-ranging compendium of workload measurement methodology, though most certainly some methods have been either missed or are under represented.

The objectives and scope of the effort, as approved by the NDB, were as follows:

OBJECTIVES: Military aircraft are becoming increasingly complex, the associated avionics systems more sophisticated, and the mission profiles more demanding. The objective of the Working Group is to study if such an increase in aircrew workload has become a limiting factor in the operational employment of some aircraft and to determine valuable methods to evaluate this workload.

SCOPE OF WORK: *The measurement* domain will be broken down into sensory threshold function tests, motor function, and responses to psycho, physio, and chemical excitation.

The methodology will include a wide range of instrumentation, laboratories, inflight measurement and modelling methods, with the goal of compiling systematically and evaluating the multiplicity of approaches and techniques implied.

A companion document, AGARD Advisory Report 139 (AR-139) gives the conclusions drawn by the Working Group within the bounds of the above objectives and scope.

The members of the Working Group were:

PANEL MEMBERS

R. Auffret, FR
E.P. Beck, UK
B.O. Hartman, US (Chairman)
A.N. Nicholson, UK
G. Rotondo, IT
P. Woodward, US

NON-PANEL MEMBERS

K.A. Kimball, US
M. Lees, US
G.S. Malecki, US
R.D. O'Donnell, US
F.S. Pettyjohn, US
A. Roscoe, UK
M.G. Sanders, US
H.M. Wegmann, GE

Numerous other members of the Aerospace Medical Panel attended Working Group meetings, because of the high level of interest in this topic within the panel.

LIST OF CONTRIBUTORS

Richard A. Albanese
USAF School of Aerospace Medicine
Brooks AFB, Texas, 78235, USA

Jackson Beatty
Department of Psychology
University of California at Los Angeles
Los Angeles, California, 90024

Clyde A. Brictson
Anthony P. Ciavarelli
Dunlap and Associated, Inc.
Western Division
7765 Girard Ave., Suite 04
La Jolla, California, 92037

Edward P. Buckley
William F. O'Connor
Tom Beebe
Federal Aviation Agency
National Aviation Facilities Experimental Center
Atlantic City, New Jersey, 08405

R. Cannings
Royal Air Force Institute of Aviation Medicine
Farnborough GU14 6SZ
Hants, UK

Billy M. Crawford
Systems Research Branch
Human Engineering Division
Wright-Patterson AFB, Ohio, 45433

Walter B. Gartner
Miles R. Murphy
333 Ravens Wood
Menlo Park, California, 94025

G.H. Lawrence
Office of Naval Research
Physiology Program (Code 441)
800 North Quincy Street – Room 433
Arlington, Virginia, 22217

Richard E. McKenzie
Edward P. Buckley
Kiriako Sarlanis
Federal Aviation Agency
National Aviation Facilities Experimental Center
Atlantic City, New Jersey, 08405

Richard E. McKenzie
Bryce O. Hartman
Crew Technology Division
USAF School of Aerospace Medicine (AFSC)
Brooks AFB, Texas, 78235, USA

Carl E. Melton
Aviation Physiology Laboratory
Civil Aeromedical Institute
FAA Aeronautical Center
Oklahoma City, Oklahoma, 73125

Layne P. Perelli
Crew Technology Division
USAF School of Aerospace Medicine (AFSC)
Brooks AFB, Texas, 78235, USA

Alan H. Roscoe
Royal Aircraft Establishment
Bedford, UK

Prof. Gaetano Rotondo
Italian Air Force Medical Service HQ
Via P. Gobetti 2 – 00185 Rome, Italy

R. Simmons
M. Sanders
K. Kimball
US Army Aeromedical Research Laboratory
Fort Rucker, Alabama, 36362, USA

Walter W. Wierwille
Robert C. Williges
Virginia Polytechnic Institute and State University
Blacksburg, Virginia, 24060, USA

Samuel G. Schiflett
Naval Air Test Center
Patuxent River, Maryland, 20670, USA

CONTENTS

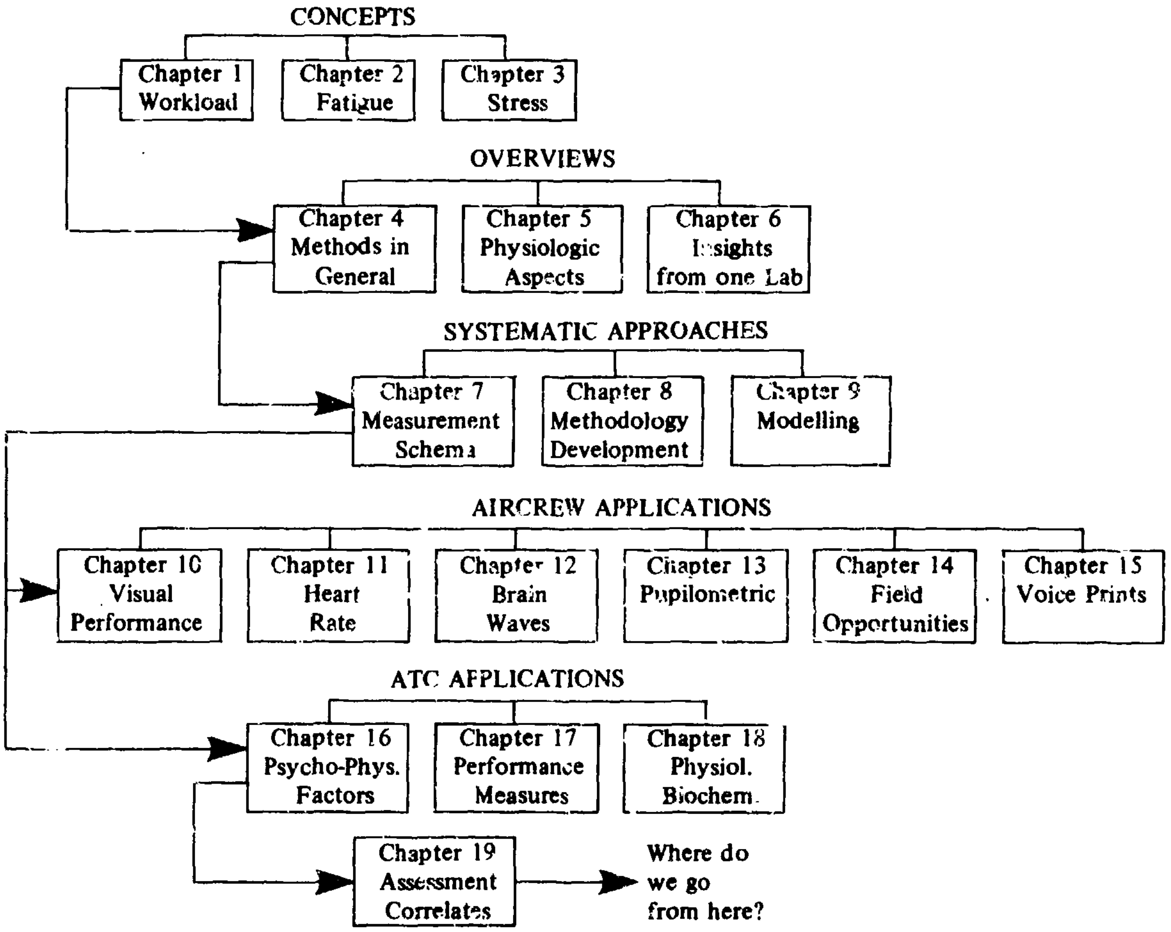
INTRODUCTION

Task complexity is everywhere in the environment within the operational pilot functions: avionics systems, commonly with a digital computer core and a wide range of sensors and information displays; a cockpit packed with flight displays and controls; capabilities and, at times, requirements for multiple missions, confrontation with a variety of threat systems; crowded airspace; multiple command/control/target designation systems and techniques; a host of environmental burdens, inside and outside the cockpit. In addition, the NATO nations have seen the emergence of multi-role aircraft and an expansion in the tactical employment of the helicopter. One result of these technological and operational advances has been a marked increase in aircrew workload. This increase in workload has become a problem of operational significance, to the point where, in some cases, aircrew capability has become a limiting factor in the operational employment of some aircraft in the more demanding missions. As a consequence, problems of aircrew workload have assumed increasing importance in the NATO research community.

Methods of measuring workload have a substantial history in the NATO research community. Disciplines represented include systems design engineering, operations research, the behavioral sciences, aerospace medicine, physiology, biochemistry, and biotechnology in general. There has been considerable variation in the kinds of experimental tasks employed, the measures obtained, the instrumentation used, the analytic models and methods employed, the ratio of synthetic modelling versus empirical data used, and the kinds of laboratory facilities required. The measurement domains include measures of sensory threshold, measures of sensory integration, cognitive function tests, measures of motor function, vigilance, reaction time, psychophysiologic responses, physiologic and biochemical changes. Methodology includes a wide range of instrumentation, laboratory facilities and environments, inflight measurement methods, and modelling methods. Analysis models and experimental design requirements also vary considerably. Computer utilization in the areas of experimental programming and data processing has become commonplace. Periodic overviews of current findings are necessary. There is a need for summary matrices, as well as a widely endorsed taxonomy of human performance.

This AGARDograph is one such periodic overview. It is current in the sense that each chapter is a condensation or modificant of recent papers, prepared specifically by each author to fit the objectives of this Working Group. Ongoing research involving advances in workload measurement technology obviously cannot be represented in this report, since the editors avoided tasking contributors with the preparation of chapters "de novo." Such is the nature of "periodic overviews."

It will be helpful to the reader to have a "road-map" of this report. Diagrammatically, it looks like this:



CONCEPTS OF WORKLOAD*

by

Walter B. Gartner, Ph.D.
Miles R. Murphy, Ph.D.
333 Ravens Wood
Menlo Park California 94025

In ordinary uncritical discourse, the phenomena referred to by the terms "pilot workload" and "fatigue" are easily distinguished. In its broadest and simplest aspect, pilot workload refers to how much a pilot must do to perform a specified flight operation. Fatigue is widely understood as a feeling of tension or weariness, often accompanied by an obvious unwillingness or inability to continue to work or perform. However, when attempts are made to quantify the workload imposed on a pilot by a particular aircraft design, or operational procedure, or to access the effects of fatigue upon system performance, important unresolved issues arise in regard to the more precise specification of workload and fatigue concepts and to the adequacy of assessment criteria and techniques. This chapter and the next address the principle unresolved issues in conceptualizing and measuring pilot workload and fatigue. In a survey of the origins of operator workload concepts, Jahns (1) has found it useful to characterize workload as "an integrative concept for evaluating the effects on the human operator associated with multiple stresses occuring within man-machine environments." Further, he proposed to partition this broad conception of workload under three functionally related components: (1) input load, (2) operator effort, and (3) work result.

While broader conceptions may be considered useful for indicating the range and diversity of workload reference, the purpose here is to outline the principle ways in which investigators have elected to restrict the use of the term. Therefore, we will discuss Jahns' basic classification scheme with only some minor changes in terminology.

Workload as a Set of Task Demands: The common attribute of task-demand concepts of workload is the use of the term to refer to requirements for task performance which can be specified without reference to any operator response or activity actually applied to satisfy these requirements. The distinction between demands, as such, and any actual operator response (including capabilities, readiness to respond, etc., is a very important one. One approach to the treatment of workload as demand is exemplified by Klein's (2) attempt to quantify and predict "design-specific instantaneous workload levels imposed upon the pilot while in flight." In distinguishing this approach from traditional workload quantification methods, Klein emphasized that "workload is addressed from the standpoint of predicting human performance requirements as demanded by the system and its operational environment rather than from the standpoint of measurement of human responses to those demands."

The application of task analysis techniques within a designated system-mission-environment context to determine task performance requirements is a familiar and widely used human factors practice. Gartner, et. al, (3) proposed that demands be more strictly defined as inputs to the crew, which actually serve (directly or indirectly), to establish crew performance objectives or to represent operational conditions and events which in an actual flight situation would be expected to initiate crew activity or modify ongoing crew responses. Task demands are identified using functional criteria, that is, they are inputs that operate as response programs or as action requirements for the crew. This distinction between response and stimulus-oriented expressions of task demand is considered to underlie some of the problems in workload assessment and practical application, because different kinds of demands are often confounded. In other words, system-oriented and situation-specific demands are often confused with perceived demands by the operator or with the behaviorial or psychophysiological demands imposed on an operator by an assigned task. This task-demand concept is closely related to Jahns' (1) input load component which he defines operationally as "a vector (L) of input data which must be transformed by the operator into a vector (P) of output data to satisfy a specified performance criterion function and/or maintain a homeostatic operator state." This input load characterization of task demand fits a variety of operator-loading concepts that distinguish one or more sensory channnels or modalities as important to task performance, and addresses such concerns as channel capacity, perceptual overload, and so forth. For example, in his review of task load factors, Hartman (4) defined load as "the sum of all requirements imposed on the operator at any instant by the system," and later distinguished load as the number of information channels affecting operator performance.

The defining feature of demand oriented expressions of workload is simply that they be free of any dependence upon considerations of operator response or response capabilities. In view of the apparent difficulty in sustaining this distinction in practice, it is probably advisable to associate task demand only with input or stumulus-oriented variables and to reserve workload for the response-oriented variables.

Workload as Effort: The focus of the conceptualization of workload as effort relates to how much a operator has to do, and/or how he must work to satisfy a specified set of demands. A general characterization of this concept of workload has been given by Cooper and Harper (5): 'The term workload is intended to convey the amount of effort and attention, both physical and mental that the pilot must provide to attain a given level of performance."

A somewhat different emphasis is provided by Welford (6) in characterizing effort as "the intensity with which action is carried out. A man may work either more or less hard at a job." Here, the emphasis shifts from effort required to the consideration of the effort a human operator actually does exert in the performance of a task.

---

* This chapter was abstracted by the editor from NASA TN D-8365, Pilot workload and fatigue: a critical survey of concepts and assessment techniques with permission of the authors.

In his elaboration of the operator-effort component of workload, Jahns emphasizes the operator's readiness to respond and he identifies such factors as experience, motivation, set, physiological readiness and physical factors, as well as the general background and personality of the operator as determinants of this operator's state.

The concept of effort is most often used simply to refoer to how hard a man is working and not to the actual task performance or to the difficulty or demands of the task. Singleton (7) has argued for the separation of performance and effort by invoking the familiar observation that "an operator may be performing better in one of two tasks as compared in an experiment because he is trying harder rather than because one task is easier than the other." Whatever it is that occurs when a man is working harder is referred to as effort.

Workload as Activity or Accomplishment: The conceptualization of workload as activity to actual task performance or the products of this activity. It is often used in operational studies of the effects of operating procedures or system design on aircrew performance. In a summary report (8) on a UAL-ALPA joint project to evaluate pilot workload the authors (who are not named) defined workload in terms of the total activity of the captain and co-pilot in performing such tasks as flight-path control, vigilance, communications, navigation, and system operation during each phase of a actual flight. The actual activities engaged in by crew members have also been used as workload referents in long-term studies of crew performance factors. For example, Cantrell and Hartman (9) recorded typical flight-crew activities over 20 consecutive days, including off-duty and administrative activities as well as those carried out in flight, to be used as an index of workload.

Workload Assessment Techniques: A critical review of workload assessment techniques Gartner and Murphy (10) indicates that dispite conceptual and practical difficulties the attempt to develop and apply useful measures of pilot workload is being vigorously pursued. The workload techniques which they examined included task-demand analysis, measures of task performance, psychophysiological measures and subjective reports. None of these assessment techniques were found to be free of significant limitations in their sensitivity to differences in task difficulty, in distinguishing between physical and mental effort, or in the reliability of data acquisition and interpretation procedures.

With respect to workload, Gartner and Murphy recommend that significant improvements in both measurement and management can best be accomplished by refinements and innovations in the analysis and measurement of pilot effort. They state, "human-factors engineering activities are already being applied to task-demand analysis, and effective techniques are available for this application." However, systematic attempts to assess effort *per se* are considerable less in evidence, despite the fact that such assessments are needed for the empirical evaluation and adjustment of task demands. Innovations in the direct assessment of effort would also provide a basis for developing more effective "effort control" techniques. They also point out that there are directly assessible neuromuscular tension patterns which can be reliably related to both central neurophysiological states and the task-relevant phenomena of attention and perception.

In summary then, it can be seen that there are several ways of conceptualizing workload, though in general they might be divided into an emphasis on the input side (task demands) or the output side (the work output). Similarly, there are variations in the appropriate measurement techniques though here we see no obvious simplification. The diversity of definitions and approaches accounts for this working group (AMP WG-08) report, and is a condition which should be kept in mind as the reader proceeds thru this document.

REFERENCES

1. Jahns, D. W. Operator Workload: What is it and how should it be measured? Crew System Design. Proceed of an Interagency Conference on Management and Technology in the Crew System Design Process, Los Angeles CA., September 12-14, 1972.

2. Klein, T. J. A workload simulation model for predicting human performance requirements in the pilot-aircraft environment. Paper presented, Human Factors Society's 14th Annual Convention, San Francisco, CA., October 13-16, 1970.

3. Gartner, W. B., Erenetz, W. J. and Donohue, V.R. A full mission simulation scenario in support of SST crew factors research, NASA, CR-2150, 1972.

4. Hartman, B. O. Time and load factors in astronaut proficiency. Symposium on Psychiophysiological Aspects of Space Flight, B. E. Flakerty, ed., Columbia Univ. Press, NY., 1961.

5. Cooper, G. E. and Harper, R. P. Jr. The use of pilot rating in the evaluation of aircraft handling qualities, NASA, TN D-5153, 1969.

6. Welford, A. T. Fundamentals of skill, Methuen and Co., London, 1968.

7. Singleton, W. T. Prologue to Section 1: Measurement of man at work, W. T. Singleton, J. G. Fox and D. Whitfield, eds., Taylor and Francis, London, 1971.

8. Anon. Some workload and environmental characteristics of an air carrier short haul turbo-jet operation. Summary reports on a UAL-ALPA joint project to evaluate pilot workload on B-737 flight operations, United Airlines, 1969.

9. Cantrell, G. K. and Hartman, B. O. Application of time and workload analysis technics to transport flyers, SAM-TR-67-71, AFSC, Brooks AFB, TX., April 1967.

10. Gartner, W. B. and Murphy, M. R. Pilot workload and fatigue: a critical survey of concepts and assesment technqiues, NASA TN D-8365, November 1976.

CONCEPTS OF FATIGUE*

by

Walter B. Gartner, Ph.D.
Miles R. Murphy, Ph.D.
333 Ravens Wood
Menlo Park California 94025

A good summary statement of the recurring theme that the central difficulty in dealing effectively with the problem of fatigue is one of definition was presented by Welford (1). According to him, fatigue means a subjective state following some kind of physical or mental strain in ordinary "man-in-the-street" constructs. However, to the physiologist, fatigue means some kind of reduction of response following more or less prolonged activity. However, the psychologist, is placed in the middle and charged with the responsibility of tackling the problem of fatigue relative to practical human affairs. Unfortunately, according to Welford, he often evades this responsibility by dismissing fatigue as unscientific or by redefining the phenomena.

In another reference, Welford (2) notes that, "difficulties have led some wish to abandon the term fatigue, yet there is a need for a term to cover those changes in performance which take place over a period of time during which some part of the mechanism, whether sensory, central, or muscular, becomes chronically overloaded." Bartley (3) develops the position that the inherent utility of the concept will be realized only when it is clearly distinguished from such considerations as: (1) situation in which it occurs, (2) the bodily expression of fatigue, and (3) the effects of fatigue on performance, work output, and so forth. However, it will be apparent in the following overview of fatigue concepts that such phenomenon have not been excluded for more restrictive definitions of fatigue, and that considerable diversity in the contemporary use of the term remains. Part of the problem seems to be related to the wide overlap between the concepts of workload and those of fatigue. In the detailed overview of these concepts, Gartner and Murphy (4) demonstrate that within an average of 30.5 workload and fatigue indicators, well over 50% of the indicators are directly or indirectly related or overlapped to a significant if not indistinguishable degree.

Fatigue as a Feeling of Weariness or Tiredness: This conceptualization of fatigue has been characterized by Bartley (3) as experimental or sensory-cogitative. Experimental concepts seem to be favored in operational studies of fatigue wherein extensive use is made of subjective assessments. In his review of operational studies, Schreuder (5) elaborates on the subjective aspects of fatigue to suggest that, "The ordinary sense of weariness which the pilot subjectively feels after a hard day's work should not be labeled as fatigue." Schreuder would insist on a level of intensity of this feeling of weariness "which is an excess of the expected normal fatigue and which is cumulative and of such amount as to alter the pilot's judgement and ability." Factor analytic studies of fatigue indicate that the sensation of fatigue has three major components: (1) bodily tiredness and drowsiness, (2) weakened motivation or concentration and, (3) a group of physical complaints, not unlike those of psychosomatic disorders. Other investigators are satisfied with more global and unqualified definitions: Yoshitake (6) "The feeling of fatigue signifies overall unpleasantness experienced by workers and is not quite the same as complaints of symptoms of fatigue."

Fatigue as a Clinical Syndrome: In clinical practice, subjective complaints and/or specific sets of signs and symptoms are regarded as useful working definitions for fatigue. Mohler (7) has outlined an extensive list of signs and symptoms for both physical and mental fatigue, with the physical signs expressed primarily in terms of physiological functions, i.e., increased blood glucose, increased lag in pupillary response, instability of neuromuscular coordination, etc. Mohler's mental symptoms are expressed in terms of psychogenic and emotional dysfunction and include increased irritability and intolerance, tendency to depression and withdrawal, and increased sex drive, etc.

Hartman (8) suggests a three-category classification of fatigue (acute, cumulative and chronic) characterizing acute fatigue as that normally occurring between a pair of sleep periods, and cumulative fatigue as occurring over a period of day or weeks as a result of inadequate recovery from successive periods of acute fatigue. Hartman urges a clinical definition of chronic fatigue as "a psychoneurotic syndrome characterized by difficulty in committing oneself to a active or aggressive course of action, and by a generalized withdrawal or retreat from conflict which is intolerable for situational or personality reasons."

Fatigue as Performance Decrement or Skill Impairment: Fatigue concept referents in this category, like the clinical signs and symptoms just cited, are often treated as indicators or effects of fatigue rather than a distinguishable state. For example, Bartlett (9) states "Fatigue is a term to cover all those determinable changes in the expression of an activity which can be traced to the continuing exercise of that activity under its normal operating conditions, and which can be shown to lead, either to deterioration in the expression of that activity, or more simply, to results within the activity that are not wanted."

A more formal expression of these changes in performance is provided by Hull's development of the reactive-inhibition construct (10). Hull's behavioral restatement of Spearman's general law of fatigue and Pavlov's concept of conditioned inhibition is: "Whenever any reaction is evoked in an organism there is left a condition or state which acts as a primary, negative in that it as an innate capacity to produce a ceasation of the activity which produced the state, we shall call this state or condition reactive inhibition".

---

* This chapter was abstracted by the editor from NASA TN D-8365, Pilot workload and fatigue: a critical survey of concepts and assessment technqiues with permission of the authors.

One of the more interesting variants is Bartlett's (11) concept of "skill fatigue." On the basis, of studies pilot performance in the Cambridge psychological laboratory he suggests that it is, "Necessary to draw a broad distinction between fatigue produced by continued hard physical work and that produced by work which calls for little continuous muscular effort, but demands persistent concentration and a high degree of skill." Skill-fatigue, also distinguished from mental fatigue, is said to occur when a task, such as piloting a plane, requires complex, coordinated, and accurately timed activities. In other Cambridge studies, deterioration of skill performance was apparent after about 2½ to 3 hours of stimulated flying,manifesting primarily as a progressive lowering of standards of performance, the missing important information displays, and the gross mistiming of interrelated control actions.

Fatigue as a Neurophysiological Condition or State: In traditional or classical studies, fatigue was referred to a particular neuromuscular site, that is, to specific motor units or muscle groups or organs or tissue structures and then defined in terms of specific biochemical and/or response capability changes. This comparatively narrow focus is now generally recognized as only one aspect of fatigue. In a discussion of neuromuscular fatigue as a special instance of a more general condition as Basmajian (12) points out "I shall observe, at once the traditional and necessary warning that fatigue is a complex phenomenon and perhaps a complex of numerous phenomenon. The fatigue of strenuous effort is probably quite different from the weariness felt after a long day's routine sedentary work. Undoubtedly, the following types exist: emotional fatigue, central nervous system fatigue, general fatigue, and perpherial neuromuscular fatigue of special kinds."

Welford (13) feels that fatigue is best conceptualized as a local neural impairment.

While Grandjean (14) shares the view of many investigators that fatigue is a central neurophysiological condition and is located in the central nervous system more specifically, in the brain stem reticular activation system. His conceptualization of fatigue as a conditon of the central nervous system is based on early studies of the role of the brain-stem reticular formation in producing and maintaining various levels of inactivity, arousal, and activation.

Welford has also suggested that considering fatigue as a central phenomenon attempts to integrate the comparatively less accessible condition of mental fatigue with the more readily observed condition of neuromuscular fatigue. He states, "It appears that in the intact organism changes in the muscles brought about by prolonged or repeated contractions can, according to circumstances, have one of two limiting effects. Either the muscles themselves become temporarily incapable of further contraction or the condition of the muscles produces afferent stimuli and these in term affect the central mechanisms and lead to the ceasation of efferent impulses." If the term mental fatigue is to have a meaning in line with that of neuromuscular fatigue, it must denote the impairment of some brain mechanism as a result of long continued use. The impairment must be reversible in the sense that it disappears with rest, and may take the form of lowered sensitivity, or lowered responsiveness, or lowered capacity.

This definition by Welford permits a distinction to be made between mental fatigue and other central conditions such as adaptation, habituation, and monotony or boredom which also lead to a decrement in performance over time. However, others see no significance differences in operational definitions of reactive inhibition, habituation, and central fatigue. Grandjean (13) expresses the popular view that boredom are components of the fatigue condition and are related to the task situation: "If the workload is too heavy, fatigue due to physical or mental effort is to be expected; if the worker is underloaded or forced to conduct repetitive work, fatigue due to monotony will be produced."

Fatigue as a Level of Energy Expenditure: The energy expenditure approach to fatigue focuses on the cost of protracted effort, whether mental or physical, in terms of the energy investments or transformations required to sustain it. A formulation of the energistic approach by Dukes-Dobos (15) defines fatigue as a term to denote a normal psychophysiological process which starts immediately after the beginning of any physical or mental activity and which consists of the utilization of the bodies' energy stores, the accumulation of the breakdown products, and the activation of adaptive mechanisms which maintain the homeostatis of the organism."

Cameron (16) considers the term of fatigue to be no more than a useful descriptive term for a generalized stress response over a period of time. "The human stress response is generalized in character, involving the whole system of biological emergency mechanisms. Since it implies, by definition, an abnormal demand on the energy resources of the system, it is fatiguing. The degree of fatigue experienced may depend to some extent on the level of the stress response, but will depend primarily on its duration." Here, he emphasizes the duration of the stress response of the organism, not necessarily the duration of the stressful condition. This is a critical distinction, because he argues that the length of time needed to return to a normal arousal level, that is a normal level of biological emergency mechanism activity is good index of the severity of fatigue.

McFarland (17) has criticized the focus on physiological factors and fatigue citing the familiar arguments that effects observed in the laboratory are not always found in actual work situations and that other factors often influence energy reserves and utilization capacities, mainly physical condition and motivation, and that the metabolic costs of mental work are very slight. In this argument, characterizations of the pilot's job as predominantly cognative and not physical or muscular are frequently cited to question the relevance of physiological factors, especially those derived from studies of heavy physical workloads. It would seem that if this concept of fatigue as a level of energy expenditure is to bear fruit we must have a clear focus on the higher order concepts of energy mobilization and channeling in the individual, rather than focusing on the localization and reduction of this response to metabolic activities in particular muscles or tissues.

Welford, who views both mental and neuromusclar fatigue as effects of loading, would agree that fatigue is a consequence or concomitant of workload. Bartley would also agree with this relationship while insisting that fatigue is a condition of the individual and is not to be defined in terms of external situations or even work products. In this situation, he considers energy expenditure, paced performance, prolonged activity, and demands upon particular body mechanisms to be typically fatigue-producing.

The primary difficulty in applying fatigue assessment techniques more explicitly is the multidimensional character of fatigue phenomena and their interaction with even more complex phenomena of individual motivation and stress tolerance. The approach of Bartlett to fatigue assessment utilizing the application of the concept of skill fatigue is important since observable changes in pilot behavior during primary task performance can be clearly and directly related to the accomplishment of flight management and/or aircraft control objectives. It must be concluded that factors other than task demands or protracted effort are more significant in the occurrence of fatigue. These other factors include individual differences in personality, motivation, physical fittness, and life style, as well as such situational factors as operational management policies, disruption of established biorhythms, sleep patterns, and exposure to various environmental stressors. The relative contribution of personal versus task-specific fatigue factors is an important unresolved issue.

REFERENCES

1. Welford, A. T. The psychologist's problem in measuring fatigue. Symposium on Fatigue, W. F. Floyd and A. T. Welford, eds., H. R. Lewis and Co., London, 1953.

2. Welford, A. T. Fundamentals of Skill, Methuen and Co., London, 1968.

3. Bartley, S. H. Fatigue: Mechanisms and Management, Charles C. Thomas, Springfield IL., 1965.

4. Gartner, W. B. and Murphy, M. R. Pilot workload and fatigue: a critical survey of concepts and assessment techniques, NASA TN D-8365, November 1976.

5. Schreuder, O. B. Medical aspects of aircraft pilot fatigue with special reference to the commercial jet pilot, Aerosp. Med. Special Report Pub. 37, (4) 1966.

6. Yoshitake, H. Relations between the symptoms and the feelings of fatigue. Methodology in Human Fatigue Assessment, Taylor and Francis, London, 1971.

7. Mohler, S. R. Fatigue in aviation activities, Aerosp. Med., Vol, 37, 1966.

8. Hartman, B. O. Psychological factors in flying fatigue. Psychiatry in Aerospace Medicine, Vol. 4, C. Perry, ed., Little Brown and Co., Boston MA., 1967.

9. Bartlett, F. C. Fatigue in the air pilot, Rept. 488, Air Ministry, Flying Personnel Research Committee, 1942.

10. Hull, C. L. Principles of Behavior, Appleton-Century-Crofts, NY., 1943.

11. Bartlett, F. C. Fatigue following highly skilled work, Proc. Roy. Soc. Ser. B. Vol. 131, 1943.

12. Basmajian, J. V. Muscles alive: their function revealed by electromyography, The Williams and Williams Co., Baltimore, 1974.

13. Welford, A. T. Fundamentals of Skill, Methuen and Co., London, 1968.

14. Grandjean, E. Introductory Remarks. Methodology in Human Fatigue Assessment, Taylor and Francis, London, 1971.

15. Dukes-Dobos, F. N. Fatigue from the point of view of urinary metabolites. Methodology in Human Fatigue Assessment, Taylor and Francis, London, 1971.

16. Cameron, C. A theory of fatigue, Ergonomics, Vol. 16, 1976.

17. McFarland, R. A. Understanding fatigue in modern life. Methodology in Human Fatigue Assessment, Taylor and Francis, London, 1971.

CONCEPTS OF STRESS

by

Richard E. McKenzie, Ph.D.
Crew Technology Division
USAF School of Aerospace Medicine (AFSC)
Brooks Air Force Base, Texas 78235
USA

To paraphrase the Webster Dictionary definition of stress, we find that stress is a physical, chemical, or emotional factor to which an individual fails to make a satisfactory adaption and which causes physiologic tensions that may be a contributory cause of disease. While this publication is not the format for a clinical discussion of the psychological, psychiatric, and biological aspects of the effects of stress on the airmen and its related effects upon the man-machine interface, the concept of stress and its decrementing role is an important one in modern aerospace flight.

In his discussion of psychology and flying fatigue, Hartman (1) defines acute fatigue as that which occurs in a single flight, during a single day, or more appropriately, between a pair of sleep periods. Here the recovery from acute fatigue is a function of the adequate amount of rest available. But even without prolonged or even relatively short rest periods, the fatigued flier can mobilize his resources and return briefly to near rested levels of efficiency when the occasion demands. Hartman also defines cumulative fatigue as that which occurs over a period of days or weeks, and is the result of inadequate recovery from successive periods of acute fatigue. Recovery from cumulative fatigue is also dependent upon adequate rest. However, without an adequate recovery schedule, the pilot finds himself fighting an enhanced workload self-generated by his own loss in airmanship and efficiency, and finds that the longer cumulative fatigue continues to build up, the longer it will take for him to recover his reserve and his capacity to mobilize himself to meet high demand situations.

The term "chronic fatigue" has a special psychiatric meaning and is defined as a neuropsychiatric disease. Chronic fatigue is a psychoneurotic symptom characterized primarily by difficulty in committing oneself to an active or aggressive course of action and by a generalized withdrawal or retreat from a conflict which is intolerable for situational or personality reasons. Thus, this entity is rarely seen in either the military or civilian pilot/aircrew member.

Like fatigue, stress has its acute phases, one of which is an alerting arousal response enabling the person to perform better and to otherwise adapt himself to an emergency. Cumulative stress, on the other hand, is a build-up of physiological, chemical and emotional factors over a period time until some kind of maladaption occurs. As Selye (2) points out, stress is a reasonably normal component of modern every day life and can be adaptive, but cumulative stress becomes maladaptive and ultimately then, stress becomes distress.

Selye (3) has also discussed his general adaptation syndrome which he conceptualizes as the defensive response of the body, through the endocrine system, to systemic injury evoked by stress. This is worked out by an initial stage of shock, like an arousal or surprise reaction, followed by a stage of growing resistance to the injury (adaptation), followed in turn by a final stage of healing or exhaustion and death if adaptation fails. Note that there is no alternative course of action, one must either resist the bodily effects of stress by healing or one must become exhausted, and ultimately be defeated by the effects of stress. In short, then, one cannot ignore the effects of cumulative stress.

Sparks (4) in a chapter entitled "The Clinical Aspects of Psychiatric Illness in Fliers," brings to light a relatively interesting aspect of a pilot's career progress. One could call this situation one of selective screening, because flying personnel, through the almost automatic process by which they learn the skills required for modern pilotage, be it military or civil aviation, are screened for emotional stability. During their initial period of training, they have close supervision, with exposure to moderate levels of stress together with the requirements of rigid discipline demanded by attention to procedures and the awareness that the aerospace environment is an unforgiving mistress. In order to successfully pursue his chosen career, the pilot must adapt to the early stresses of flight training. Flight itself poses additional stressors to which he must form adaptive methodologies or strategies which act as a further screening process. Following the completion of flying training, additional periods of flying duties, upgrading and so forth, further cause him to adapt more and more strategies for coping to the point that almost any unstable individual would be self-eliminated prior to any operation assignment. For the military pilot, combat poses additional and unique stressors to which he must adapt, ground himself, or be ultimately surrendered. It is a small wonder that we expect to find any psychiatric casualties once this screening process is completed.

However, the very aspect of recognizing this screening process implies that we also recognize that we have a highly selected individual who will almost invariably stand up to most of life's stresses. Therefore, we are inclined to form an almost mythical concept of the pilot as being inviolate and unaffected by ordinary and inordinate stressors. This of course, is far from true. Carlos Perry in a chapter on aerospace psychiatry (5), examines some the stressors of aerospace operation. He points out that "potential danger, physical discomfort, energy demands, attitude, and enforced physical passivity have been well recognized as stresses and need little further elaboration. He also shows that increased specialization in aerospace operations is the source of stressors that were not apparent when flying activities were more generally uniform. Thus, a given airman may well tolerate stresses associated with flying long conventional type cargo hauls in the company of a crew, but not be able to successfully cope with stresses of solitary, short duration, high altitude intercept flights in high performance single engine jetcraft. He points out the incongruity involved in the military concept of alert. Here, we have aircrew men who are interested in flying, going places, and seeing things and we force them to sit for long periods of time in an alert facility, away from family and other satisfying activities. Perry states out that boredom has also become a major stress factor. Here automation, lack of diversification, endless routines, and increasing length of individual flights contribute to the production of boredom. He states that "while boredom may be considered to be a benign type of stress, that these feelings are not far removed from the more serious feelings of lack of ambition, futility, or even depression." He points out that the nature of aerospace operations are source of another major category of stress, with the necessities of long and frequent travel, varying periods of absences which can be a source of severe stress to maritial and parental activities.

Even the complexities posed by the various types of aircrew equipment for providing a livable environment for man imposes their own constraints and physiologic stresses. Mission and operational requirements present the modern pilot and crew with everchanging complex tasks which provide another form of stress. These major sources of aircrew stress are compounded by the individual's internal psychophysiologic reaction to stress and to general external stressors, such as personal, career or family problems.

The human body is known to adapt to or to withstand severe conditions of immediate or acute stress. However, it does not respond as well to long-term or cumulative stress, whatever its source. While aircrew personnel are subject to special forms of stress the basic reaction to stress is uniform. All stress, physical, emotional and so forth, is responded to by some kind of an adaptive or avoidance reaction. The basic need is to protect oneself from more and more stress. Physically the initial response is a musculoskeletal tension/arousal response, which together with the corresponding changes in glandular, organ, and nervous systems prepare the individual to retreat from the situation or to comfront it, the classic flight or flight reaction.

In the case of cumulative stress, the musculoskeletal and organ system of the body tend to be continually activiated to the point where the individual is now stressed even when the original source of stress is absent. The stress becomes internalized and most stimuli, either internal or external, become sources of stress. We now have a chronically tense, irritable, agitated, disturbed person.

As cumulative stress continues, the musculoskeletal and organ systems of the body may start to undergo pathologic changes. We begin to see psychosomatic systems of stress in the muscles of the neck and shoulders or other parts of the body. Chronic muscle tension produces decreased blood flow in these tissues, with pain and joint pathology. Chronic organ reactions yield typical symptoms of the gastrointestional tract, such as stomach cramps, ulcer, colitis, and so forth.

It is obvious that chronic stress and its related pathology cannot be ignored. The aircrew member who experiences cumulative stress from one or multiple sources and who is then further subjected to additional increments of stress from personal or aircraft equipment, from the demands of the mission, or from fatigue or external stressors will find his best skills and efforts decremented. This degrading of performance is obviously related to the disaster or near-disaster of the aircraft accident, but not necessarily in the causal sense. A recent survey of USAF accidents fails to support degraded performance or stress as a causative agent. Instead, stress and decremented performance are seen as factors which are contributory in that they act to "set the stage", preparing the psychologic and physiologic world of the pilot in such a manner that he is not able to respond effectively to one or more additional untoward events. This is the insidious danger of stress pathology and constitutes an excellent reason for including a pre-accident mental status investigation as part of the accident review process. However, as in most things, "an ounce of prevention is worth a pound of cure." How can we prevent stress, or better yet, how can we prevent dis stress as the result of cumulative stress?

There is a solution for cumulative stress. It has long been known, in a simple minded way, that one cannot be tense and relaxed at the same time. Thus, the adaptive response to stress/tension is one of relaxation. Adequate recovery times from periods of cumulative stress with provision for recreation are important. Perhaps even more important, however, would be a conditioned learning program wherein the individual is taught to avoid the effects of cumulative stress by keeping himself in a relatively relaxed state. There are several long-time approaches to this type of training. One of the earliest being that of Schultz with his autogenic training followed by Jacobson's progressive relaxation training and more recently by such mediative techniques as transendental mediation (TM). An intriguing modern day addition to these forms of relaxation methodologies is that of biofeedback techniques where an electronically generated signal from the muscle or organ system involved is available to the individual as a learning technique. This is based on an axiom in information theory that states "the controllor giving information about the state of the system can then exercise control over that system." It has been demonstrated that the central nervous system can exercise exquisite control over the CNS, the spinothalamic ard automonic nervous systems. Biofeedback is merely one way of giving the controller information about the state of the system so that he can learn to exercise the necessary control. While biofeedback simply utilizes modern electronic technology, we must realize that there are many adaptive strategies which could be employed and also realize that human beings come supplied with internal biofeedback signals which undoubtedly play an important role in both adaptive and maladaptive behaviors.

REFERENCES

1. Hartman, B. O. Psychological factors in flying fatigue, Psychiatry in Aerospace Medicine, J. G. Perry, (ed.), Little, Brown and Company, Boston, 1967.

2. Selye, H. Stress Without Distress, J. B. Lippincott Co., 1974.

3. Selye, H. Stress, Acta Inc. Medical Publishers, Montreal, 1950.

4. Sparks, J. C. Clinical aspects of psychiatric illness in flyers, Psychiatry in Aerospace Medicine, J. G. Perry, (ed.), Little, Brown and Company, Boston, 1967.

5. Perry, C. J. G. Aerospace psychiatry. Aerosp. Med., 2nd. Ed., Hugh W. Randel, (ed.), The Williams & Williams Company, Baltimore, 1971.

6. McKenzie, R. E. and Baines, R. H. Biofeedback therapy. In: What's psychotherapy and who needs it. (ed) by H. L. Collier, O'Sullivan, Woodside and Company. Phoenix, AZ., 1977.

SOME CONSIDERATIONS CONCERNING METHODS TO EVALUATE AND ASSESS WORKLOAD IN AIRCRAFT PILOTS

by

Professor Gaetano Rotondo, Brigadier General IAF, MC
Italian Air Force Medical Service H.Q. - Italy
Via P. Gobetti 2 - 00185 Roma

If the nature and entity are analyzed of the various stressing and fatiguing factors that are acting on the body and psyche of aircrafts' pilots during their specific activity, it appears obvious that underlying the exercise of a pilot's profession is a basic situation which ultimately permeates the whole of his activity and exerts a multiplicity of effects on the physique and psyche of the same pilot.

The fundamental characteristics of such a situation may be summarized as follows:

1. Flying involves the use of a machine that is required, unlike other machines, to respect certain aerodynamic laws. Any infraction of these laws involves an immediate risk of crash and accident. In the pilot's professional activity, therefore, life depends on the machine and its continuous efficiency, a situation that in actual practice expresses itself in the form of a permanent image of potential "vulnerability" undoubtedly present in the subconscious of each and every pilot.

2. Pilot's activity depends a great deal on the spatial environment of the aircraft, the three-dimensional displacement and rapid transition of the aircraft, and, indirectly, on the various conditions that have repercussions on the human body (i.e., accelerations, acoustic and non-acoustic vibrations, equipment, sensorial stress, etc.). All these conditions constitute links in a chain of factors which readily explain the wealth of interferences that act on the somato-psychic balance and, consequently, on performance, adaptability and, in the long run, on individual fatigue.

3. Flying does not just represent a technical or operative activity, i.e., a job, but rather "a vital activity and an 'in toto' reaction of the ego to the environment."

Upon such a basic substrate, which is in itself potentially stressing and qualitatively common to all pilots irrespective of their specialization and the type of aircraft they fly, there then act interferences due to the various physical and psychic factors, each of which plays a specific and individualizing role both in connection with general and particular aircraft (fighter, transport, reconnaissance, rescue, or helicopter, etc.).

It would certainly be interesting and important if it were possible to define the degree and limits of such psychophysical workload by means of technically valid and acceptable scientific methods with a view to obtaining differential qualitative and quantitative assessments of the various flying specializations. In fact, numerous methods have been proposed periodically for obtaining a measure of workload by quantitatively evaluating the functional changes that fatigue can produce. As known, such changes may consist of an increase of the duration and inconstancy of the psychomotorial reaction times; an increase of the latency time of the pupillar reflex; a diminution of the capacity for rapid binocular fusion; an increase of the accommodation time for near and distant vision; a reduction of the critical flicker fusion frequency (3, 12) and changes of other ophthalmic indexes; modifications of the characters and duration of the monosynaptic spinal reflexes produced, for example, in the area of the sciatic nerve (1); variations in the duration of the central nervous time of the orbicular blinking reflex under light stimulation, and the time needed for a complex mental process (11); reduction in muscular force and muscular tone; increased instability in neuromuscular coordination; increased loss of electrolytes through cutaneous sweating; reduced circulating plasma volume; variations in the urinary excretion of corticosteroids (7) and cathecolamine (2, 6); variations in the lactacidemia, glycemia, and cholesterolemia values, the ratio between alpha and beta lipoproteins, the number of the eosinophiles, and the hematocrit index; and, finally, electrocardiographic changes and variations in the Ruffier and Dickson index of cardiac resistance (4).

Quite obviously, however, all these methods lend themselves very readily to criticism. Indeed, none of the results obtained by these methods are capable of being interpreted in a unique manner. In fact, these methods measure functional changes that are or can be influenced considerably by a wealth of other factors, both endogenous and exogenous, including first and foremost, the subject's age. Therefore, if one wanted to make a comparative evaluation of the amount and precocity of the stress and the psychophysical workload produced by the individual stressing factors connected with flying, one would admit that it is extremely difficult to find a precise differential criterion that could be used to obtain a quantitative graduation of this workload. This is not only because the subjective element, here understood as the individuality and extreme variability of the response of the single subject to every type of stimulus, has a predominant weight in this particular activity; it also depends on the nature and entity of the reaction to any kind of stimulus which are, in turn, conditioned by numerous and extremely variable individual, environmental, and circumstantial factors.

After those necessary premises concerning the difficulties of a unique interpretation of all proposed diagnostic methods and the preponderance of psychical workload on physical one in the piloting aircrafts, it is our opinion that between the above-mentioned functional changes eventually produced by emotional and psychic fatigue, a particular attention could be reserved - in Aviation Medicine - to variations in the urinary excretion of corticosteroids and especially catecholamine.

It is well known that every stress - no matter if physiological or emotional - is capable of inducing organic reactions due to the increase of corticosteroids and catecholamines in the blood circulation. According to many authors who have studied the phenomenon in the aviation field, there are increases also in particular flight conditions, particularly those likely to set up a state of stress.

Therefore, it is possible to conclude that the determination of the urinary excretion of catecholamines in particular, as indication of a possible psychic stress, could be used as a method to objectify "emotions." This could have a useful application in practice to reveal emotional states undergone in flight, particularly during the phase of training and other all conditions of considerable psychical engagement in the course of aeronavigation.

In other words, the determination of such substances would then give useful information about the presence of stress and would also allow to evaluate the intensity of the latter (and of consequent workload). The same evaluation may also be obtained by determining the quantity of vanilmandelic acid (VMA) excreted with urine, such an acid taking its origin from the metabolism of catecholamines (5).

These methods might be usefully and practically employed with the purpose of obtaining an objective measurement of the emotional aspects of the human personality in real conditions, and then quantitatively evaluating the workload (especially psychic, but also physical and physiological workload) in the pilot's professional activity.

REFERENCES

1. Gualtierotti, T., R. Margaria, and D. Spinelli. 1958. Effects of stress on lower neuron activity. Exper. Med. Surg. 16:166.

2. Dlepping, J., O. Buisson, J. Guerrin, A. Escousse, and J. P. Didier. 1963. Evaluation de l'elimination urinaire des catecholamines chez des pilotes d'avions a reaction. C.R. Soc. Biol. 157:1727.

3. Krugman, H. E. 1947. CFF as a function of anxiety reaction: an exploratory study. Psychosom. Med. 9:269.

4. Le Roux, R. 1960. La fatigue operationelle des pilotes d'helicopteres. Revue des Corps de Sante 1:493.

5. Paolucci, G., and G. Blundo. 1973. Determination of emotional condition in student pilots during air-navigation by dosing vanilmandelic acid (VMA) excreted with urine. Riv. Med. Aer. Spaz. 36:184.

6. Paolucci, G., and G. Blundo. 1975. Catecholaminic excretion in student pilots. Riv. Med. Aer. Spaz. 38:27.

7. Rotondo, G. 1955. On the treatment of pilots affected by operational fatigue with dehydroisoandrosterone. Riv. Med. Aer. 18:78.

8. Rotondo, G., and A. M. De Angelis. 1966 Acetil-aspartic acid and citrulline in treatment and prevention of flight fatigue. Riv. Med. Aer. Spaz. 29:85.

9. Rotondo, G. 1969. Experimental contribution to preventive and therapeutic treatment of flight fatigue. Riv. Med. Aer. Spaz. 32:231.

10. Rotondo, G. 1977. Workload and operational fatigue in helicopter pilots. Aviat. Space Environmental Med. (in print).

11. Spinelli, D., and P. Cerretelli. 1961. Analysis of central nervous functions in particular physiological conditions. Med. Sport. 1:128.

12. Vozza, R. 1955. Flicker fusion frequency as a test of operational fatigue in jet pilots. Riv. Med. Aer. 18:771.

PHYSIOLOGIC ASPECTS OF WORKLOAD/FATIGUE/STRESS*

by

Layne P. Perelli, Captain, USAF
Crew Technology Division
USAF School of Aerospace Medicine (AFSC)
Brooks Air Force Base, Texas 78235
USA

It is important to recognize that the physiological mechanisms of the organism do not particularly care nor are they necessarily aware that they are reacting to the effects of workload, the effects of fatigue or the effects of stress. Physiological mechanisms provide a common link between the concepts of workload, fatigue and stress. Traditionally the basic physiological approach to fatigue involves the measurement of energy expended in performing a given amount of work. As early as 1919 to 1920, Waller and De Decker (1) measured the carbon dioxide production of workers and were able to relate increases in carbon dioxide production to a reduction of work output during a night's activity. They use the term "physiological cost" to describe the increased metabolic demands resulting from increased fatigue and related lowered performance. Page (2) (3) has suggested that the concept of fatigue be replaced with the concept of metabolic cost, and Bitterman (4) has suggested that the concept of fatigue be defined as a reduced efficiency resulting from continued work and reversible by rest, with efficiency defined as the ratio of performance to expended effort. Effort was to be determined from metabolic cost indices.

Concepts of physiologic cost are related to Selye's concept of the general adaptation syndrome in which any stress to which the body is exposed creates an overall non-specific, systemic reaction to cope with or reduce the stress (5). It is theorized that fatigue creates a stressful condition to which the body tries to adapt, and in so doing produces an abnormal set of physiologic indicators which can be evaluated as to the severity of the fatigue/stressor.

After reviewing several fatigue studies showing no significiant or dramatic performance decrement and one study with a performance increase, Cameron (6) concludes that performance measures are too erratic and unreliable to serve as indicators of fatigue. He feels that the term fatigue should be used as no more than a descriptive term for generalized stress response over a period of time, and that the best index of acute and chronic effects would be the time required for biologic emergency mechanisms to return to a normal arousal level.

Pursuing this same line of thinking, Harris and O'Hanlon (7) provide a review of what is known about the recovery of man from exposure to certain adverse conditions, such as sleep deprivation, abnormal work/rest cycles, prolonged physical work, and environmental and situational stressors. Their purpose was to determine if recovery functions can predict how long a man can maintain effective performance before he must be relieved and how long a rest period is required before he is ready again to perform effectively during continuous military operations. They conclude that while there is insufficient knowledge now available to make such predictions, the following list of potential physiological failures seems most important to consider and reversal of these impairments may provide practical indications that recovery has taken place: 1) Degraded physical working capacity, 2) Inadequate iron reclamation, 3) Mild cardial fatigue, 4) Paroxysmal cerebral cortical activity, 5) Impaired carbohydrate metabolism, 6) Thiamine deficiency, 7) Involuntary hypohydration, 8) Glycogen exhaustion, 9) Increased susceptibility to infection, 10) Imbalanced protein metabolism and 11) Adrenal cortical and medullary exhaustion. They feel like Cameron that changes due to fatigue will become apparent in the physiological systems before performance degradation occurs. This implies, of course, that even though a given schedule of work has not yet produced performance decrement, work-rest cycles should be structured so that severe changes in the physiological systems are prevented.

Following this same concept that the physiologic cost of fatigue is generally not an immediate problem providing the individual receives sufficient recovery time, Hartman and Cantrell (8) have taken the position that the best approach to maintain man's capacity for skillful work is to engineer the system so that physiological degradation is eliminated. This implies that if physiological indicators known to be associated with stress reactions are found to be within normal limits, then it is presumed that no performance decrement of operational consequence has occured. Thus, the problem is to quantify these physiologic limits in relation to a criteria of performance degradation in such a way as to cause system managers to design, man, and use an operational system in such a way that these limits are not exceeded. One of the difficulities in using physiological indicators for evaluating workload, fatigue, or stress is of a temporal nature in that some physiological responses can be observed only after periods of hours or even days while other responses occur almost instantaneously. Some measures are unobtrusive which could be used in operational situations while others are somewhat impractical or often impossible to obtain.

We will first discuss the long term physiological indicators of stress workload and fatigue recovered from the organism and measured as urinary metabolites, namely the 17-hydroxycortico steroids (17-OHCS) and the catecholamines (epinephrine and norepinephrine). The following general review of 17-cortico steroids and catecholamines is taken from Guyton (9).

Steroids, namely cortisol are excreted into the blood stream from the adrenal cortex in response to a wide variety of stresses. Steroids enable the body to cope with stress through its effects on carbohydrate, fat and protein metabolism. It causes a stimulation of gluconeogenesis by the liver and a decrease in glucose utilization by the cells which in turn raises the blood glucose concentration. At

---

* This material was abstracted from a chapter of Captain Perelli's draft doctoral dissertation by Richard E. McKenzie, Ph.D.

the same time it causes a reduction in protein stores in all parts of the body except the liver. Blood amino acid concentration goes up, transport of amino acids into extra hepatic cells is diminished and transport of amino acids to the liver is enhanced. Amino acids are thus mobilized from the tissues to the liver. Finally, fatty acids are brought out of adipose tissue increasing their blood concentration which increases their utilization for energy. The adrenal cortex secretes steroids in response to adrenocorticotrophic hormones from the adenohypophysis which is under direct control of the hypothalmus. With this indirect feedback mechanism, levels of cortisol can continue to rise to very high blood concentrations as long as the stress agent continues to stimulate the hypothalmus in some way. Cortisol fixes to its target tissues in about 20 minutes after release. The normal blood concentration is about 12 micrograms per 100 milliliters and its half life in the blood is 100 minutes. The normal secretory rate is 15 milligrams per day of which approximately 75% is excreted in the urine.

At this point it should be obvious that one can measure 17 keto-steroid production from either blood or urine sampling. The only problem one should be aware of is that there is a difference in the concentration time of 17-OHCS found in blood plasma as opposed to urine by about two hours. Increases in 17-OHCS excretion have been found for various anxiety producing situations, such as electroshock treatment and with the use or administration of hallucinogenic drugs and in the viewing of mildly stressful motion pictures. Berkum, Bialek, Kern and Yagi (10) performed an extensive series of experiements simulating five stressful military situations in which the subject was led to believe that he was in immediate danger of losing his life or of being seriously injured, or that by his actions he has seriously injured one of his colleagues. All of these stress situations resulted in elevated 17-OHCS excretion and the level of the increase was related to the presumed level of stress induced for each situation.

Miller (11) provides a review of the many studies in which 17 keto-steroids have been found to increase due to the stress of military flying. In 1943, Pincus and Hoagland (12) conducted three sets of experiments which related steroid excretion and flying stress. They reported not only significantly increased steroid production, but that individual performance scores were positively related to the level of steroid increase. They also reported that increases in steroid production were found to be related to independent rating given by the pilot's squadron commander on their individual susceptability to fatigue.

Catecholamines: Catecholamines are secreted by the adrenal medulla in response to stimulation from the sympathetic nervous system. The relation-ship between the adrenal medulla and a threating situation was first demonstrated by Cannon and de la Paz (13). While the proportions of catecholamines which are excreted depend upon the physiologic conditions, on the average, 75% epinephrine and 25% norepinephrine are excreted. Their effects on the body are the same as those caused by direct stimulation of the sympathetic nervous system, but the effects last about ten times longer since the circulating catecholamines are only slowly removed from the blood. It should be noted that the sympathetic nerve endings excrete norepinephrine, but in a matter of seconds it is reabsorbed or destroyed at the cellular level by O-methyl transferase or monoamine oxidase. These enzymes are similar to cholinesterase which destroys acetycholine, the agent excreted by the parasympathetic nervous system. While both the sympathetic nervous system and the excretions of the adrenal medulla have general nonspecific effects, the catecholamines stimulate and increase the metabolic rate of every cell in the body. However, it must be noted that circulating catecholamines do not readily pass the blood-brain barrier. This means that central nervous system physiology is not as reactive to these circulating substances as is the rest of the body physiology.

The general result of stimulation of the sympathetic nervous system is to mobilze the body for action. Norepinephrine causes general vasoconstriction, increased cardiac activity, increased basal metabolism, sweating, inhibition of the gastrointestinal tract, glucose release from the liver, decreased kidney output, and adrenocortical secretion. Epinephrine has similar effects but has a greater stimulating effect on cardiac activity and basal metabolism and has a less constricting effect on the vascular system of the skeletal muscular system. Normal resting secretion rates are .2 micrograms per kilogram of body weight per minute for epinephrine and .07 micrograms per kilogram of body weight per minute for norepinephrine.

While there is some indication that catecholamines are excreted due to stress, they are generally released in relation to the overall activity level or performance level. In a review of catecholamine response to various activities, Euler (14) reports that mental stress associated with anger, aggression, or exhilaration will increase norepinephrine excretion while emotional states characterized by apprehension, discomfort, painful or unpleasant feelings, will increase epinephrine excretion. As an example of what one may expect to find in measures of catecholamine levels, Euler and Lundberg (15) found that urinary epinephrine levels were elevated in pilots as well as inexperienced passengers during one or one and one-half hours of moderately stressful flights. The pilots also had elevated norepinephrine levels while the passengers did not. Melton and Fiorica (16) found that both epinephrine and norephrine excretions were elevated during cross-country flights in private pilots with less than 100 hours flying experience. However, the levels of excretion were not related to the length of flying time. A more recent study by Krahenbuhl, Marett and King (17) explored catecholamine production during various phases of Air Force flying training in the T-37 jet aircraft. They found that the emergency procedure phase which was given in a Link trainer was essentially nonstressful, but that both epinephrine and norepinephrine were significiantly elevated from control values during actual spin, solo and check flights. Here again, the assessment using epinephrine appears to be more responsive than the use of norepinephrine as an indicator.

Even though there does not appear to be any functional relationship between the adrenal medulla and the adrenal cortex, there is an interaction of catecholamine in steroids effects within the body. Broverman, Klaiber and Vogel (18) have attempted to differentiate the effects of short-term versus long-term stress relative to the interaction of catecholamine and steroids. Short-term stress is hypothesized to facilitate performance on serially repetitive, overlearned tasks and to impair performance on novel tasks requiring perceptual restructuring. Long-term stress is hypothesized to have the opposite effects. They attempt to account for these finding by argueing that during short-term stress behavior is dominated by the sympathetic nervous system. However, with increasing exposure of the central nervous system to the stress-elicited adrenal hormones, dominance shifts to the parasympathetic system causing an overall depression of activity.

The Cardiac Indicators: The cardiac activity indicators heart rate (HR) and heart rate variability (HRV) have been used extensively to analyze inflight pilot activity probably because the data can be collected without gross interference of flight activities. In addition, heart rate can be measured for specific segments of performance during relatively short time spans. There is no way to precisely determine the relative contributions of any segment of behavior during a urine collection period and thus, urine analysis is confined to relatively gross estimates of when performance decrement has occurred. In addition, heart rate and heart rate variance appears to be more closely related to activity levels and performance quality than does information on catecholamine production revealed by urine analysis. One other advantage of heart rate activity is that data reduction can be almost immediately and easily performed while urine analysis requires one or two days of chemical analysis in the laboratory under fairly optimal conditions. A study by Bateman, et. al., (19) shows that heart rate for commercial pilots on routine flights, upgrade training flights, and simulator flights are very similar and higher than resting rates. However, basic training flights were found to be significantly higher. Heart rate increased in response to specific inflight stresses and when pilots were demonstrating maneuvers requiring a high degree of skill. Opmeer and Krol (20) found that increases in heart rate and decrease in heart rate variance matched the predicted order of increasing difficulty of four phases of flight, namely baseline, level flight, take-off, and approach. When pilots were required to fly realistic flight plans in a simulator, the same relative changes were found. They found heart rate variance to be a more sensitive measure than heart rate alone and they concluded that heart rate variance appeared to be more related to cognitive tasks where heart rate was more responsive to anxiety inducing tasks.

Roscoe (21) has demonstrated that heart rate is a useful tool in evaluating pilot workload changes created by new aircraft instrumentation and advanced control systems. Heart rate was found to vary as changes in weather conditions and different runways created more stressful landings. While these inflight cardiac indicators have yielded some information on cognitive workload and stress levels experienced by pilots, laboratory studies in which the stimulus presentations can be more precisely controlled have been much more successful in relating these indices to performance and workload.

The normal resting heart rate exhibits a relatively large degree of beat to beat irregularity (HRV) referred to as sinus arrythmia. Ettema and Zielhuis (22) found that sinus arrythmia was significantly depressed and heart rate, blood pressure and respiration rate were significantly increased as workload increased. They concluded that this effect is due to a change in both the breathing pattern and a rise in vagal tone and sympathetic nervous activity induced by the mental load. Boyce (23) found essentially the same increase for heart rate and decrease in HRV for increasing mental loads. A series of studies by Thackray (24) has shown HRV to be a useful measure for separating rest peroids from mental work periods on a variety of tasks. Using a two dimensional compensatory pursuit tracking task, he found that heart rate variance along with heart rate, blink rate, respiration rate, respiration period variability, and skin conductance were all capable of differentiating the rest period from the work periods. In a simulated radar control task, heart rate variance was found to be higher for subjects reporting high boredom. In addition, the performance of the subjects in the higher boredom group also significantly declined over the rest period. This would suggest that HRV reflects a level of attentiveness which is related to overall performance capability.

A fairly comprehensive view of the relationship between cardiac indicators and performance has been stated in the broader framework of arousal theory. It is known that the level of performance quality is related to the degree of arousal or activation level of the operator in terms of an inverted, U-shaped function which implies that an optimal level of activation will produce maximum performance capability. This in turn is related to the reticular activating system which in effect mediates the sleep/wakefullness dimension. This of course is related to increasing levels of fatigue. Heart rate can be expected to decrease as the subject's level of arousal falls or to increase as extra effort is put forth to stay awake. The seemingly paradoxical increase in heart rate with fatigue is normally seen with physical exertion as well, where heart rate continues to increase under vigorous exercise up to the point of the collapse of the organism. Thus, the task demands of the systems operator job must be taken into account if one is to predict the arousal level of a long duration flight. Corcoran (25) attempted to separate the concept of arousal from task demand by requiring minimal activity from subjects during a 60-hour period without sleep. In this case, both heart rate and performance on an unarousing, nonphysical, 30-minute vigilance task fell consistently. He argues that performance will follow the inverted "U" previously described with decreasing arousal, and that arousal will fall with lack of sleep or increased fatigue, but the effort to remain awake which is what is being measured by physiological indicators will be a function of task demand and subjective motivation to remain awake.

Extrapolating from these research findings, the following changes in heart rate and heart rate variances can be predicted for long duration flights. First, heart rate and heart rate variance would tend to increase with moderate levels of fatigue. With very high levels of fatigue, heart rate would be expected to fall and heart rate variance to increase still further. We would also find that tasks which created greater levels of arousal because of their complexity or the amount of concentration required would be initially more resistant to fatigue effects. From this we can hypothesize that straight and level periods of flight requiring minimal control input and instrument monitoring should show greater performance decrement with fatigue than periods when maneuvers must be performed. Heart rate should be higher and heart rate variance should be lower as the arousal value of the task increases. Tasks requiring maximum levels of information and concentration should show least performance decrement and greatest heart rate increases and greatest heart rate variance decreases.

Thus, we appear to be at a point where the important pilotage aspects of information processing, decision making, pattern recognition and so forth, are the important task variables and cardiac indicators are one of the important measures of workload, fatigue, and stress relative to the man-machine system. However, it would be a mistake to focus upon single physiologic variables. We have the present capability to collect and evaluate multiple physiologic variables and weigh them by means of regression analysis so as to investigate whether meaningful physiologic profiles can identify specific reactions to specific aspects of workload, fatigue or stress.

While we are considering the present state of the art for both ground-based and inflight physiologic measurements, some exciting breakthroughs are on the horizon which may allow us to measure and utilize cortical indicators of dynamic brain activities including decision making and information processing. But first, we shall explore how we arrived at the position that information processing activities relative to required information input is a vital consideration in the evaluation of the man-machine interface.

REFERENCES

1. Waller, A.D. and De Decker, G. E. The physiological cost of work in various departments of "The Times" printing house, J. of Physiol., 1919-9120, 53, vc-cvi.

2. Page, R. M. On supplanting the industrial fatigue concept, J. of Business, 1929, 2, pp. 137-153.

3. Page, R. M. Measuring human energy cost in industry: a general guide to the literature, Genetic Psycho. Momo, 1932, 11, pp. 321-537.

4. Bitterman, M. E. Fatigue defined a reduced efficiency, Amer. J. of Psychol., 1944, 57, pp. 569-573.

5. Selye, H. The physiology and pathology of exposure to stress, Montreal: Acta, Inc., 1950.

6. Cameron, C. A theory of fatigue, Ergonomics, 1973, 16, pp. 633-648.

7. Harris, W. and O'Hanlon, J. F. A study of recovery functions in man (U.S. Army Tech. Memorandum 10-72), Aberdeen Proving Ground, MD: Aberdeen Research and Development Center, Human Engineering Laboratory, April, 1972 (NTIS No. AD-741 828).

8. Hartman, B. O. and Cantrell, G. K. Psychomotor monitoring for general efficiency, In Medical Education for National Defense Symposium on Biomedical Monitoring, Brooks AFB, Texas: USAF School of Aerospace Medicine, November 1964.

9. Guyton, A. C. Textbook of Medical Physiology (4th Ed.), Philadelphia: W. B. Saunders, 1971.

10. Berkum, M. M., Bialek, H. M. Kern, R. P. and Yagi, K. Experimental studies of psychological stress in man, Psychol. Mono., 1962, 76, pp. 1-39.

11. Miller, R. G. Secretion of 17-hydroxycortico steroids (17-OHCS) in military aviators as an index of response to stress: a review, Aerosp. Med, 1968, 39, pp. 498-501.

12. Pincus, G. and Hoagland, H. Steroid excretion in the stress of flying, J. of Aviation Med., 1943, 14, pp. 173-193.

13. Cannon, W. B. and delaPaz, D. Emotional stimulation of adrenal secretion, Amer. J. of Physiol., 1911, 27, pp. 64-70.

14. Euler, U. S. Quantitation of stress by catecholamine analysis, Clinical Pharmacology and Therapeutics, 1964, 5, pp. 398-404.

15. Euler, U. S. and Lundberg, U. Effect of flying on the epinephrine excretion in Air Force personnel, J. of Appl. Physiol., 1953, 6, pp. 551-555.

16. Melton, C. E. and Fiorica, V. Physiological responses of low-time private pilots to cross-country flying (FAA-AM-71-23), Oklahoma City, Oklahoma: FAA Civil Aeromedical Institutes, April 1971.

17. Krahenbuhl, G. S., Marett, J. R. and King, N. W. Catecholamine excretion in T-37 flight training, Aviat, Space and Environ. Medicine, 1977, 48, pp. 405-408.

18. Broverman, R., Klaiber, E. L., Voge, W. and Kobayashi, Y. Short-term vs. Long-term effects of adrenal hormones on behaviors, Psychol, Bull., 1974, 81, pp. 672-694.

19. Bateman, S. C., Goldsmith, R., Jackson, K. F., Ruffell Smith, H. P. and Mottodes, V. S. Heart rate of training captains engaged in different activities, Aerosp. Med., 1970, 41, pp. 425-429.

20. Opmeer, C. H. J. M. and Krol, J. P. Towards an objective assessment of cockpit workload: 1. physiological variables during different flight phases, Aerosp. Med., 1973, 44(5), pp. 527-532.

21. Roscoe, A. H. Use of pilot heart rate measurement in flight evaluation, Clin. Med., 1976, 47(1), pp. 86-90.

22. Ettema, J. H. and Zielhuis, R. L. Physiological parameters of mental load, Ergonomics, 1971, 14, pp. 137-144.

23. Boyce, P. R. Sinus arrhythmia as a measure of mental load, Ergonomics, 1974, 17, pp. 177-183.

24. Thackray, R. I. Patterns of physiological activity accompanying performance on a perceptual-motor task (AM 69-8), Oklahoma City, Oklahoma: FAA, Office of Aviation Medicine, Civil Aeromedical Institute, August 1975.

25. Corcoran, D. W. J. Changes in heart rate and performance as a result of loss of sleep, Brit. J. of Psychol., 1964, 55, pp. 307-314.

SOME INSIGHTS RELATIVE TO THE MAN-MACHINE SYSTEM:
AN OVERVIEW OF TEN YEARS OF RESEARCH

by

Richard E. McKenzie, Ph.D., Bryce O. Hartman, Ph.D.
Crew Technology Division
USAF School of Aerospace Medicine (AFSC)
Brooks Air Force Base, Texas 78235
USA

It is known that man-machine systems require certain kinds of operator skills and involve specific kinds of tasks whether they are ground based, air-borne or in space. Viewing the development of aviation from its infancy through current operational aircraft, airborne weapons systems, and space systems we see a remarkable accelerated development of automation. With this development, there has been a shift in the nature of the job performed by the man in this assembly of man and machine. In general, piloting is really more like "machinemanship," with the number of subsystems which the pilot must control, the number of cockpit displays and other informational inputs as well as the increased communications load, all contributing to a tremendous increase in workload (1). The term workload is a somewhat ambiguous concept that can be defined in many ways. We feel that workload encompasses the concepts of performance, fatigue, and stress, any one of which can be defined in terms of the other. Keeping in mind the pilot's function as a systems monitor, wherein he initiates occasional commands to the system, we know that the pilot will assume actual control of the system only at intervals against a background of activity at a lower level. Thus, we have a highly variable work rate situation and our initial concern is whether or not the intervals of low activity might alter the efficiency of the operator when he is required to assume command or exercise control over the system. In our first attack on this problem we used four different workload levels from which the subjects went into a period of overload. The subjects in this study were used in a single session, matched group design. There were a total of 20 subjects, five in each of 4 load levels. We obtained pronounced decrements in performance during overload after successively lower work load levels. Unfortunately, in spite of the matching there were some differences between groups of subjects on initial or baseline proficiency, therefore, we felt that this initial exploratory study was not an adequate evaluation of the problem. What we needed was a repeated session design using each subject as his own control. With this refinement in a follow-on experiment, we found no differences in proficiency related to different base work rates. This confirmed British studies on speed, that is signal rate stress, wherein the effects obtained are function only of the <u>immediate</u> operator load and are independent of the characteristic of the preceding task load levels. So we found that the system operator works at a steady systematic rate independent of the more variable rate of signal onset. The operator tends to ignore a rapid onset of signals, preceding in a methodic fashion to work on each subtask as he gets to it. We liken this smoothing function to the strategy of "queing" proposed by Miller. In this strategy the operator assigns each new input to a kind of conceptual list of responses to be made when he gets to them. We looked for Miller's other adaptive strategies which he called "filtering" (ignoring some signals in order to process the remaining more effectively) and "two-handed operation." Instances of filtering could not be identified and two-handed operation occured only infrequently; however, this does pose the question as to what conditions cause or promote the use of such strategies (2).

This initial study was reported in 1961 but in the meanwhile we became involved in evaluating system operator performance factors in the School of Aerospace Medicine's space cabin simulator. In evaluating the operator data, we reported the possible "energizing" effect of an intial high signal rate period had on a subsequent period of very low signal rates. We also felt that signal rate might be a way of manipulating both duty time and diurnal variables (3). With the idea that performance decrement was not specifically time-anchored, but more of an immediate or instantaneous product related to signal rate, we continued to gather data in the space cabin. The next series of flights explored a reversal of day/night operating times. Here we again found that signal rate was a primary factor in performance, with marked decrement at low signal rates below those of 119 per hour. This effect is attenuated by the day/night cycle in that performance decrement is not as great when the low signal rate periods occurred during the day (4).

At this point the requirement for evaluating special mission personnel including astronaut candidates led to the interesting concept of task induced stress. Here competing tasks were used in a manner so as to cause the operator to psychologically internalize the task stress, rather than to attribute his obvious performance deteriment to the task itself. Aside from the problems of selection and evaluation the results of our attempt to induce this kind of stress show that the competing task situation produced significant task stress which could be used to access the relative adaptiveness of the individual. In other words, the selected group was better able to perform and was, therefore, less susceptible to the signal/noise ambibuity produced by the task and less bothered by the induced task stress (5).

A later overview of all of the SAM space cabin flights was aimed at evaluating information input as a factor in crew performance. We might have called the study "signal rate revisited". In brief, we were able to show that a constant, fairly high level of signal rate (500 signals per hour) resulted in a rather remarkable increase in operator performance compared to low and/or variable work loads (6). Thus, we have placed work/rest cycles, diurnal variations, etc., in the role of secondary or even more remote factors in human performance, and we are left with information input as a critical variable. We can infer that any factor affecting man's ability to process information; that is, fatigue, drugs, stress (physical and psychologic), etc., will be reflected in decremented performance.

One easily recognized confounding factor in the work load/performance area is fatigue. Ordinary fatigue has never shown up as a significant factor in the space cabin studies. However, operational requirements often impose aircrew problems related to change of sleep cycles, early awakening, and so forth, in the face of increased pilotage demands. At times, tactical demands have raised the question of preflight and/or in-flight pharmacologic support. One such demand indicated the use of preflight sleep induction and inflight arousal via medications. Two research studies were designed to test the effects such medication might have on performance. Here, workload was an approximation of the tactical mission. Performance was measured using the Multi-dimensional Pursuit Test developed by the USAF School of Aerospace Medicine years ago as an aid to pilot selection. Preselection criteria and a psychologic test battery were used as predictors. Heart rate and respiration were also monitored. In this instance, physiologic monitoring and psychologic testing did not reveal nor predict any systematic changes related to the drug treatments.

The drug treatments involved the administration of secobarbital (three grains) with the "in flight" administration of d-ampethamine (5 milligrains) with appropriate controls. The results indicated a hangover effect of three grains of secobarbital seen at the start of the mission 10 hours later and still present at the end of the mission 12 hours later. The effects of d-ampethamine are decreased in individuals taking secobarbital (7).

A follow-on study using only 1½ grains of secobarbital showed no apparent psychomotor hangover (8). While these kind of in-laboratory studies are needed, the cost of doing more than approximating the task structure and workload requirements are usually prohibitive. However, we feel that the increased necessity to consider the use of pharmacologic agents by aircrew members will dictate more studies relating to pilot performance. What the laboratory lacks is some comparative standard of laboratory task or task system as it relates to acceptable performance standards for actual aircraft pilotage. In spite of the long history of laboratory testing, we still cannot answer the question "Will this particular drug, or instrument, or device impair or enhance the pilot's ability to perform his required duties?" Theoretically, it should be possible to state with scientific confidence that performance on laboratory task "X" at a certain level indicates that pilotage under the experimental conditions being evaluated would be difficult, dangerous or impossible. We have not yet developed such criteria which can be applied to the airborne human. However, we try!

Given the strong evidence of the critical nature of the relationship between information processing ability and aircrew performance, perhaps we should make a dedicated effort to evaluate information processing ability as our laboratory task "X" and compare these results with simulated aircraft pilotage performance.

REFERENCES

1. Hartman, B. O. and McKenzie, R. E. The complex behavior simulator - a device for studying psychologic problems in modern weapons systems, USAF School of Aerospace Medicine Report 61-9, December 1960.

2. Hartman, B. O. and McKenzie, R. E. Systems operator proficiency: effects of speed stress on overload performance, USAF School of Aerospace Medicine Report 61-40, June 1961.

3. McKenzie, R. E.; Hartman, B. O. and Welch, B. E. Observations in the SAM Two-Man Space Cabin Simulator: III. System operator performance factors, Aerosp. Med. 32:603, June 1961.

4. Hartman, B. O., McKenzie, R E. and Welch, B. E. Performance effects in 17-day simulated space flights, Aerosp. Med. 33:1098, 1962.

5. McKenzie, R. E. A systems task used in the stress testing of special mission personnel, Human Factors, December 1965, pp. 585-590.

6. McKenzie, R. E. Crew performance as a factor of information input, Aerosp. Med. 44:3, 1971.

7. McKenzie, R. E. and Elliott, L. L. Effects of secobarbital and d-amphetamine on performance during a simulated air mission, Aerosp, Med 36:8, 1965.

8. Hartman, B. O. and McKenzie, R. E. Hangover effect of secobarbital on simulated pilotage performance, Aerosp. Med 39:11, 1966.

AIRCREW WORKLOAD ASSESSMENT TECHNIQUES

by

Walter W. Wierwille and Robert C. Williges
Virginia Polytechnic Institute and State University
Blacksburg, Virginia USA

and

Samuel G. Schiflett
Naval Air Test Center
Patuxent River, Maryland USA

Abstract

A classification scheme is presented which summarizes a survey and analysis of aircrew workload assessment techniques relevant to inflight test and evaluation considerations. Two dimensions consisting of universal operator behaviors and workload assessment methodologies were used in the classification scheme. The universal operator behaviors were classified according to the Berliner, Angell, and Shearer (1964) categories including perceptual, mediational, communication, and motor processes; wheras the workload assessment methodologies were cataloged into 28 procedures under the general categories of subjective opinion, spare mental capacity, primary task, and physiological measures. An applicability matrix based on this classification scheme is presented which summarizes existing research on workload assessment methodologies, and a bibliography of over 400 relevant references is provided as an appendix to this paper. Procedures are described whereby this matrix can be used as a guide for selecting candidate aircrew workload assessment measures for inflight evaluation. A brief overview of the various workload assessment techniques is presented along with a set of critical criteria that need to be considered in evaluating the feasibility of these measures for in-flight environments. It was concluded that no one single technique can be recommended as the definitive measure of operator workload, but the resulting classification scheme and applicability matrix can aid the investigator in choosing among presently available techniques.

INTRODUCTION

One need only compare the cockpit of a modern jet fighter to its World War II predecessor to appreciate the dramatic increase in cockpit complexity. Technological advances during the past 30 years have resulted in sophisticated avionics and weapons delivery subsystems which are available to aid the aircrew in completing a specified mission. The ultimate mission success of today's modern fighter, however, still rests on a common factor present in its World War II counterpart. This factor is the human operator. To be an effective weapon, the modern fighter with all its advanced sensors and avionics must be compatible with the capabilities and limitations of the aircrew operator.

During the design, development, and test and evaluation of any new aircraft, care must be taken that the new system does not place unreasonable demands on the aircrew by overwhelming them with too much information and too little time to process that information. Such considerations are often characterized as assessing the mental workload of the system operator.

When one reviews the research literature pertaining to mental workload, two conclusions are readily apparent. Namely, there is no single, agreed upon definition of mental workload, and there is no single, universal metric of it. Mental workload is a theoretical construct, and as such, might best be defined operationally. Clearly, it is related to factors such as operator stress and effort, but these concepts also require operational definitions. Reising (1972) provides an excellent overview of the difficulties and complexities involved in defining and measuring workload.

Rather than provide a single definition, one must consider the various operational definitions used in measuring operator mental workload. The systems engineer, for example, may emphasize operational definitions based on time available to perform a task. Psychologists tend to emphasize the information processing aspects of mental workload and operationally define it in terms of measures related to channel capacity and residual attention. Physiologists on the other hand, emphasize considerations of operator stress and arousal.

Purpose

The impetus for this report stemmed from a selective annotated bibliography of 83 references which represent potential measurement techniques for assessment of operator workload in operational environments (Schiflett, 1976). This annotated bibliography categorized the various methods in terms of general references, system analysis, subjective techniques, psychomotor performance, information processing, physiological measures, and combined methodologies. Schiflett concluded that the majority of the methods were developed for use in the design stage of aircrew systems, thereby making them difficult and/or impractical to use in the later stages of the operational test and evaluation environment.

This project was undertaken to provide a more comprehensive survey and analysis of the presently available workload assessment methodologies and was specifically directed toward the flight test and evaluation environment.

Approach

To accomplish the goals of this project a comprehensive search of the scientific literature was conducted including books, scientific journals, technical reports, and proceedings of technical meetings. Computerized information retrieval, library searches and direct contacts with the scientific community were used to locate relevant documents. Given the large pool of potential documents obtained by these combined search procedures, it was necessary to adapt a set of general and specific criteria for inclusion of a reference in the final bibliography of over 400 references appended to this paper. Details on the search procedures as well as selection criteria are provided in Wierwille and Williges (1978).

Following the selection of the appropriate workload literature, a user-oriented classification scheme which combined workload methodology with universal aircrew behaviors was used to generate a catalog of presently available workload assessment techniques. Specifically, this paper provides a description of this classification scheme and details the use of this scheme for the selection of potential measures of workload. In addition, an overview of the resulting catalog of methodologies is presented.

CLASSIFICATION SCHEME

An important prerequisite to developing a catalog of methodologies pertinent to operator workload in a flight environment is a comprehensive classification scheme. This scheme is necessary to form a basis for selecting documents, to classify the citations listed in the bibliography, and to make possible the construction of a usable analysis catalog. Consequently, the resulting classification scheme is central to a meaningful analysis of the workload literature applicable to aircrew considerations.

One dilemma that must be resolved in developing a classification scheme is that of providing a scheme with a meaningful organization of existing workload assessment methodologies. A second dilemma centers around providing a classification human operator behaviors which are related to aircrew performance so that accurate implications can be drawn from the vast amount of workload research that was not conducted in a specific aviation-related context. To solve these dilemmas, the selected scientific literature was classified according to both the universal operator behaviors present in aircrew missions as well as the specific workload methodologies.

Universal Operator Behaviors

The range of operator behaviors and their taxonomy have been investigated for several years. These behaviors have been used to obtain an understanding of what functions an operator performs in a system and as a basis for task analysis. One widely used listing of operator behaviors was developed by Berliner, Angell, and Shearer (1964). This approach breaks operator behavior into four major processes (perceptual, mediational, communication, and motor) as shown in Table 1. These four major processes are further subdivided into seven activities and then into 47 mutually exclusive operator behaviors. Because the terms used in this scheme are orthogonal, this classification can be expected to yield good agreement among investigators in determining specific behaviors for a specific aircrew problem. Consequently, the Berliner, et al. (1964) approach was used to classify operator behaviors in this report. To facilitate referencing to this classification, a graduated numbering scheme as listed in Table 1 is used throughout.

Workload Methodologies

The second dimension of classification is the specific list of available methodologies that are potentially applicable to aircrew workload assessment. The literature on workload is so diverse that categorization on the part of the reader of this literature is almost intuitive. It is, however, important to select a categorization which groups the various workload techniques in a logical way, so that conflicts and discrepancies on workload concepts are minimized.

The taxonomy of workload methods that evolved from the documents reviewed was found to be particularly useful and logical. This listing of methodologies is presented in Table 2 along with a graduated numbering designation. Basically, the various methods are grouped into four major categories (subjective opinion, spare mental capacity, primary task assessment, physiological measures) which are further subdivided into 28 individual techniques.

Literature Classification

The resulting two-dimensional classification scheme used the numerical designations of workload methodologies given in Table 2 with a subset of the universal operator behaviors given in Table 1. Early in the classification of documents according to this two-dimensional analysis it became evident that the scientific workload literature was addressed primarily to overall human performance as compared to specific, detailed aspects of performance. Consequently, the literature reviewed could be classified only according to the four major processes and seven activities shown in Table 1 instead of the 47 mutually exclusive behaviors. Even at this less-refined level of analysis, classification of the literature according to the operator behaviors dimension appeared to be more subjective and unreliable than classification on the second dimension of various workload methodologies.

Applicability Matrix

Following the abstracting and classification of the selected documents, all the references were summarized into a two-dimensional, applicability matrix which indicated the potential use of each of the 28 workload assessment techniques across the seven universal operator behaviors. A four-point rating scale was used to represent the amount of positive research evidence supporting the potential use of each workload technique for each operator behavior. These ratings included:

0: Workload method is unsuitable for assessing workload of the operator behavior cited. No research or only negative research support.

1: Workload method is potentially suitable for assessing workload of operator behavior cited. Some contradictory evidence exists; further research is needed.

2: Workload method is suitable for assessing workload of the operator behavior cited. No contradictory evidence exists; further research is needed.

3: Workload method is suitable for assessing workload of the operator behavior cited. No contradictory evidence exists. Application is proven.

The complete applicability matrix resulting from this analysis is shown in Table 3.

It should be noted that the ratings in Table 3 are based on all the research reviewed and, as such, represent data collected in laboratory simulator, field, flight simulator, and flight test environments. This was done to provide an overview of all the available data so as to suggest potentially applicable techniques for the aircrew test and evaluation environment. Conceivably, none of the data used for a particular rating was from the flight test environment. Table 3, therefore, is not totally suggestive of overall ratings of research supporting the use of a technique in the flight test environment (research of this type is, in fact, quite limited); rather it merely suggests a potentially applicable approach. To complete the evaluations for possible selection of a workload assessment technique in the flight test area, one must carefully consider the critical criteria for selection as well as the detailed evaluation of each technique. Nevertheless, considerable judgment on the part of the authors was necessary in several cases in arriving at a rating.

SELECTING A WORKLOAD ASSESSMENT METHODOLOGY

The literature summarized in Table 3 could be used for a variety of purposes. For example, cells resulting in 0 or 1 ratings could suggest areas for additional methodological research. Of primary importance, however, is the use of the classification scheme and resulting applicability matrix as an aid in the selection of a workload assessment methodology for aircrew flight test and evaluations.

Steps in Selecting a Method

The information summarized in the applicability matrix presented in Table 3 as well as the complete catalog description of workload estimation techniques presented by Wierwille and Williges (1978) can be used as a guide in selecting a workload assessment methodology in the following six step procedure:

Step 1. Specify the aircrew problem for which mental workload is to be evaluated.

Step 2: Perform a general task analysis [illegible] specific operator behaviors.

Step 3: Using the workload method applicability matrix (Table 3), calculate workload methods weighting. Rank order the methods.

Step 4: Select the first N methods in ranking. Study each of the N methods in the workload methodology literature review.

Step 5: Select the method to be used.

Step 6: Read referenced documents and plan the workload measurement experiment.

The first step is to define the particular aircrew problems for which of the mission, and particular aircrew task. The second step is to relate the aircrew problem to the universal operator behavior dimension shown in Table 1. This may be done by examining a task analysis which uses these terms or by having the investigator directly assess which behaviors are required of the aircrew member during the task. With the completion of Step 2, the aircrew problem dimension and the operator behavior dimension have been compressed into a single dimension of specific operator behaviors which can be related to the seven universal operator behaviors of the applicability matrix (Table 3).

To aid in the completion of Steps 2 and 3, a worksheet as presented in Table 4 is useful. The investigator checks the top of the appropriate columns on the worksheet of the universal operator behaviors which are germane to the particular aircrew mission as determined by Steps 1 and 2. This essentially applies equal weightings to the various operator behaviors chosen. Alternatively, each dimension can be weighted according to the importance attributed to each operator behavior present in a particular mission. For example, searching for and receiving information (1.1), information processing (2.1), and communication processes (3.) may be the central operator behaviors in a particular mission. Rather than "checking" these three dimensions on the worksheet, the investigator determines that communication processes are perhaps twice as important to the mission as the other two. Consequently, communication processes are weighted as 2 on the worksheet, and the other two operator behaviors are weighted as 1.

In step 3, the matrix of Table 3 is used to determine the applicability rating of each workload assessment technique. This is done by entering the applicable ratings from Table 3 on the worksheet and adding the row of numbers for each workload technique. If the "check" approach is used, only the rating values from Table 3 are added on the worksheet for each row and placed in the "SUM" column of the worksheet. If a weighting approach is used, the weighting is multiplied by each applicability rating number of the corresponding row; and, subsequently, the rows are added and placed in the "SUM" column of the worksheet.

Step 4 involves the rank ordering from highest to lowest score for each workload technique. The techniques with the highest scores are then selected. These N workload techniques are the most applicable for the particular aircrew problem under study. Also as a part of Step 4, the investigator reads the workload catalog summary and the bibliography pertinent to each of the N particular techniques that had the highest scores.

It is difficult to state beforehand how large N should be. Most likely, it will be between 3 and 5 for most workload problems. However, judgment on the part of the investigator must determine the value of N.

Once the investigator has read the summary of each technique, it should be possible to select the technique that is to be used. This is Step 5. Obviously, judgment again plays a major role. More specifically, practical aspects will have to be taken into consideration. Comparative difficulty of implementation, cost of the experiment, and ability to meet space, weight, and power requirements are some of the factors involved. A feasibility matrix for the selection of workload methodologies for in-flight environments is found in Table 6.

Once the technique is selected, the investigator should obtain and read in detail the documents cited in the bibliography relating to the specific technique. This will insure that available information is used in conducting and carrying out the workload assessment experiment. Pitfalls and potential misapplications might also be avoided.

SAMPLE APPLICATION

In this section the procedure for selection of one or more workload techniques will be demonstrated by a sample problem. After a brief description of the environment associated with the sample problem, the steps of the selection procedure will be described.

Background; The SS-3 Operator's Task

The SS-3 operator's position in the P-3 aircraft is one of control and usage of the aircraft's non-sonar sensors. Several sensor systems are available to the operator, and the corresponding support equipment for them is quite complex.

The SS-3 operator communicates with the remainder of the crew using an open-line intercom that is common among the entire crew. Main communications are with the TACCO (tactical aircraft coordination officer) and the pilot; however, substantial listening to the "problem being worked" is also performed by the SS-3 operator.

The SS-3 operator is responsible for ESM (electronic support measures). ESM is essentially the passive evaluation and identification of incoming radar signals. The corresponding emitters may be ships, aircraft, or ground-based radars and may be either friendly or hostile. Ordinarily, many radar signals are impinging on the aircraft. With the aid of the aircraft's central computer and the MPD (multipurpose data display) at the SS-3 position, the operator must sort and evaluate them.

The SS-3 position also contains the MAD which is designed to provide precise location information on partly or fully submerged vessels at short ranges. The system determines anomalies in the earth's magnetic field resulting from large amounts of magnetic or paramagnetic material.

In some updated P-3 aircraft, the IRDS (infrared display system) has been added. The IRDS operates at intermediate ranges between those of the radar and the MAD. It provides a television-like raster-scanned image to the SS-operator. This sensor not only provides directional information on a target or contact, but also provides an infrared (heat sensitive) picture showing details of platforms such as superstructures, rigging, antennas, snorkels or periscopes. Positive identification of the contact or target can often be made on the basis of these details.

It is important to recognize that the ESM system, the RADAR, the IRDS, and the MAD are all tied into the aircraft's central computer and appear in one form or another on the MPD before the SS-3 operator. A large portion of the operator's workload involves updating the information, selecting modes, and performing "evaluation" operations. The operator has before him, numerous command and data entry pushbuttons as well as a trackball (usable with either hand). The trackball allows the positioning of cursors and symbols on the MPD so that specific coordinates may be inputted in what appears as an analog or digital mode.

Application of the Procedure; An Example

Step 1. Aircrew workload problem statement. The IRDS is being installed in updated P-3 aircraft. In making this addition, the mental workload of the SS-3 operator is going to be increased. It is desired, therefore, to determine the workload of the SS-3 operator both with and without the addition of IRDS.

There are two points in a tactical intercept mission where highest workloads may be presumed to occur: first, when the SS-3 operator is attempting to transition from the radar to IRDS, and second, when the operator is attempting to identify the target using the IRDS (having previously acquired it). For purposes of explanation, discussion will be limited to the second high-workload condition.

When the IRDS is in use, the SS-3 operator controls the position of the sensor, vectors the pilot to the target, and keeps the TACCO appraised of details becoming visible on the IRDS display. As soon as positive identification is made by the SS-3 operator, the TACCO is informed.

Depending on the tactical situation the SS-3 operator performs ESM duties, and keeps the MPD and aircraft computer updated on the tactical situation. Furthermore the operator listens carefully to crew communications over the intercom.

During the same time period, if the aircraft does not contain an IRDS, the SS-3 operator vectors the pilot using the radar. Responsibility for positive identification is then transferred to the cockpit crew. The ESM, MPD, computer, and communication tasks remain essentially the same for the SS-3 operator during this time period.

Step 2. Determination of operator behaviors. Since the higher workloads are likely to occur with the IRDS present in the system, and since a common workload methodology should be used for both IRDS-present and IRDS-absent cases, the IRDS-present case will be used to determine operator behaviors and weightings. In the IRDS-absent case, only slight changes would occur, having to do with target identification.

In terms of the intercom task, the universal operator behaviors category (Table 1) is 3. Communication Processes. This is weighted with an importance of 4 on a scale of 0 to 5, where 0 is "of no consequence" and 5 is "absolutely critical" to mission success. Whether the SS-3 operator is verbally vectoring the pilot or providing details on the identification to the TACCO, the intercom task is very important.

The IRDS aspect of the task involves calculations of vectoring information and visual discrimination of details in the scene. These two aspects should receive a top rating of 5 because the mission is dependent on the SS-3 operator's abilities at directing and rapid identification. The task consists of 1.2 Identifying Objects, Actions, and Events, and 2.2 Problem Solving and Decision Making. Continuous tracking would also be performed. But, because slight errors in pointing the IRDS sensor would probably not harm identification (as long as the target remained in the field of view), the behavior 4.2 Complex/Continuous Motor Processes could be given a weighting of 3.

For most situations, the MAD would not yet have come into operation in the scenario, so it would be assumed that it is not part of the task. Similarly, while the radar might be operating, it would probably only be used as a back-up (when the IRDS is operating and target acquisition has already been made).

The ESM system would continue to operate and to provide information on radar emitters in the area. Under the assumption that the P-3 is not itself under attack, the SS-3 operator would relegate ESM tasks to a lower priority. The examination of radar contracts would primarily involve 1.1 Searching for and Receiving Information, 2.1 Information Processing, and 4.1 Simple/Discrete Motor Processes. This would be given a priority rating of 2. Obviously, a much higher priority would be given to ESM (probably 5) if the aircraft were under attack.

The SS-3 operator would also be performing data input duties to the MPD and computer to the extent possible. However, these aspects would be of a bookkeeping and update nature, since primary communication would be via the intercom. Nevertheless, the operator would perform the task to the extent possible. It involves 1.1 Searching for and Receiving Information, 2.1 Information Processing, 2.2 Problem Solving and Decision Making, 4.1 Simple/Discrete Motor Processes and 4.2 Complex/Continuous Motor Processes. When the IRDS identification task is being performed, MPD and computer updating might have an importance weighting of 2.

If the highest priority weighting stated above is used for each operator behavior, the weighting would appear as shown in the first horizontal line of numbers of the worksheet for this example, as shown in Table 5.

Step 3. Workload methods weighting and rank ordering. Having obtained the necessary universal operator behavior weighting for the specific SS-3 operator workload problem, it becomes possible to compute the relative weightings of workload techniques and to rank order them. This is done by multiplying each number in the "Behavior Check ( ) or Weighting" row of Table 5 by the corresponding number in each row of Table 3. Each individual product is then entered in Table 5 in the appropriate workload methodology row and operator behavior column.

All products in each row are then added, and the sum is placed in the right hand column. The workload methodologies exhibiting the highest sums are the ones most applicable to the SS-3/IRDS problem.

Step 4. Selection of N techniques; study of the techniques. The results of the selection procedure indicate that the following six techniques (ranked by numerical score with the highest first) are the most appropriate for the SS-3 operator workload problem:

2.1.1 Task Analytic; Task Component, Time Summation

1.1 Opinion; Rating Scales

1.2 Opinion; Interviews and Questionnaires

2.2.1 Secondary Task; Arithemetic-Logic (Nonadaptive)

2.2.2 Secondary Task; Tracking (Nonadaptive)

4.1.8 Physiological; Pupillary Dilation

The initial selection of six techniques rather than some other number is arbitrary. However, techniques having scores substantially below the highest ranking score are not likely to result in accurate, reliable assessment of operator workload, because the corresponding techniques are not fully proven.

It is worth noting that small changes in the weightings of importance of the universal operator behaviors would probably not have changed the outcome of the selection procedure up to this point. Most likely the same six weightings might have resulted in a different set of techniques being selected, particularly for the fourth, fifth, and sixth ranks.

After studying the six techniques more carefully using section 4 of Wierwille and Williges (1978), it should become possible to select one or possibly two to be implemented. As a means of carrying the example through the advantages and disadvantages of the six techniques will be briefly reviewed.

The task-component, time summation technique is primarily analytical. However, it could be easily adapted to the T&E environment by having SS-3 operators perform each small segment of each task separately. These could be timed. Subsequently, time available could be determined from the mission scenario, and assessment of workload determined. The apparent drawbacks to such a technique are its complexity and the fact that SS-3 operators may be capable of performing simultaneous tasks because of their high skill level.

The two opinion techniques are clearly applicable. It is probably true that the technical training of SS-3 operators is sufficient to make them highly reliable judges of their own mental workload. The investigator would have to present and specify the problem carefully so that the operators would have a clear picture of what is expected. Because of their high level of motivation, it is probable that accurate assessment of maximum tolerable workload could be obtained.

The two secondary task techniques might also be applicable. Preference should probably be given to the arithmetic-logic task, because the operator will have his left hand in use for the trackball and his right hand in use for the IRDS controller. Introduction of yet another manual control for tracking would probably cause congestion and severe intrusion. Even the arithmetic-logic task will to some degree cause congestion because the operator is already using both hands, one foot, his voice, both ears, and his vision (with at least two displays). If at all possible, the secondary task should in some way be integrated into the present task through programming. Perhaps the ESM contacts, properly attended by the SS-3 operator, could be scored as a secondary task. Since the operator would relegate this task to a low priority anyway during the specified scenario, instructions to the operator would already be similar to his present method of operating.

The technique of pupil dilation is perhaps the least proven of the six; yet it holds promise. The SS-3 operator's station in the aircraft is already somewhat isolated. A curtain can be drawn around the open side of the station, and the side window can be blocked. Consequently, ambient lighting could be maintained constant. A small video camera could probably be installed at the upper and right hand corner of the MPD. Alternatively, a commercially available, headmounted pupillography system could be used.

It should be noted that gathering of pupil dilation information is complicated by eyelid droop when the observer becomes tired. Normally SS-3 operators are on duty for six to twelve hours. Care would therefore have to be taken to fly short missions for data taking purposes.

Step 5. Workload method selection. It is believed that any of the initial six methods could be used to assess the SS-3 operator's workload. Final selection becomes a matter of ease of implementation, costs, and other matters of feasibility as indicated later in Table 6. On the basis of these factors it is most likely that an opinion approach could be most rapidly and easily implemented. It would therefore be the recommended first choice. The task component, time summation technique, using experimentally derived task element times would provide highly quantitative results. Therefore, it would be a good second choice. However, a great deal of time and effort might have to go into the experiment and the data analysis.

Step 6. Study of documents; planning of experiment. Further study of documents referenced in the appendix should make possible the construction of an opinion technique that has all the desired attributes for the particular SS-3 problem under examination. The choice of rating scales, questionnaires, interviews or some combination thereof would have to be made.

Preliminary planning of the experiment should include a test of the technique on operators who would not participate in the later data-taking session. These operators could aid in uncovering confusion terms in questionnaires or rating scales, and in ironing out problems of terminology, instructions, and scoring.

The final experimental plan should be such that the experiment, when conducted, will yield statistically significant differences in experimental conditions if in fact there are differences. The most prized result is significant differences in workload levels. Under these conditions, definite conclusions can be drawn regarding workload.

OVERVIEW OF WORKLOAD TECHNIQUES

This section provides a brief overview of the various workload estimation techniques at the second level of classification as shown in Table 2. Each procedure is described only in terms of its theory and background, because of the brevity requirements of this paper. However, a complete description of method/apparatus, areas of applications and examples, limitations, and suggested RDT & E follow-up can be found in Wierwille and Williges (1978). To provide an overall evaluation of these various techniques, certain critical criteria must first be considered in the inflight environment.

The In-Flight Aircrew Workload Problem

Whenever an attempt is made to measure workload in-flight, a number of practical considerations become important. These considerations deal primarily with the difficulties of introducing or adding anything to the cockpit or crew station environment. These practical considerations go well beyond those involved in ground simulation and have far-reaching ramifications on workload technique assessment.

Physical space. In most aircraft, crew positions are carefully designed to take maximum advantage of the space available. This space is usually limited by airframe and other design considerations in very complex trade-offs. The introduction of any device having substantial size will compromise the efficiency of the original design. Needed controls or displays may be obscured or made inaccessible. Crew comfort might also be sacrificed by reducing the already limited freedom of movement available. Therefore, for most airborne situations, desirability and practicality of the workload measurement equipment increases as the physical size of the equipment decreases. Also, there is an upper-bound on tolerable size, which for some situations might be as little as one-eighth cubic foot. However, a study of allowable size could profitably be performed.

Obviously, the physical size consideration is not as severe in ground simulation. Usually in this case a way can be found to "fit" another device into the simulation. Since the workload measurement apparatus need not be self-contained, supporting parts can be "hung" outside the simulated crew station.

Portability and self-containment. In general, it would be desirable to assess workload using a single small package that can be easily added to the crew station. Prototype and operational aircraft usually do not include power sources and telemetry or recording equipment for optional equipment in the crew station. Furthermore, present practice would probably not permit modifications of operational aircraft for powering and recording of workload measurement. The assumption, therefore, must be made that a workload assessment system must be largely self-contained if it is to be used in-flight.

Intrusion and safety. It is well known that many methods of workload measurement tend to intrude on tasks at hand (primary tasks). An aspect of intrusion that must be considered separately because of its importance is that of safety.

Certain types of flight operations are in themselves critical. Take-off, landing, ejection, and any other type of system failure, are examples of critical operations. Two types of safety-related intrusion may possibly occur through introduction of workload measurement equipment: obstruction and distraction. Obstruction involves the problem of having an extra physical object within the space needed to deal with a critical operation. Distraction pertains to the fact that the workload assessment may draw the crew member's attention away from the critical situation. Unless backup crew stations are available, it may be inadvisable to assess workload of certain critical operations in flight except by *a posteriori* techniques which by their nature do not intrude.

Data transmission or recording. It is one problem to design a feasible workload task for in-flight use--it is yet another to score the task and analyze the results. There appear to be three alternatives in the area of data analysis:

1. Perform in-flight analysis and record the processed data output in concise form for later use;

2. Record or store the unprocessed data for later playback and analysis; and

3. Telemeter or otherwise transmit unprocessed data to a ground station for recording or processing.

Experimental controls. A problem that may arise when performing in-flight experiments is that of obtaining adequate experimenter controls. The investigator or experimenter may not be on board when the workload assessment procedures are conducted. Consequently, radio contact may have to suffice. In those cases where the experimenter remains on the ground, workload assessment should be obtained by a system that is procedurally simple to operate. Also this system should be as "fudge-proof" as possible so that the effects of biases of the aircrew members are minimized.

Workload assessment integration. Because some modern aircraft incorporate computer graphic displays with substantial computer capability, the possibility exits that certain workload assessment techniques may be integrated into crew stations through software. Existing capabilities or near future capabilities may be such as to permit special modes of operation of standard displays and controls that would permit workload assessment. Scoring might be accomplished by the on board computer and the results stored in condensed form for post-flight readout. Not all methods of workload assessment may be suitable for this kind of integration; but initially, it appears that certain ones would be applicable. A feasibility study of the programming potential of new aircraft systems for workload measurement appears to be a fruitful area of research.

In-flight workload assessment summary. In-flight measurement of workload represents a challenge well beyond that of ground simulation. Factors such as physical size, weight, intrusion related to safety, portability, and experimental control become extremely important. Techniques that work well on the ground may therefore prove infeasible for in-flight use, particularly during critical mission phases such as take-off, landing, or subsystem failure (degraded mode). Nevertheless, newer techniques are becoming available that can eliminate or at least minimize the in-flight problems, principally the inclusion of workload assessment as a software change in the aircraft's avionics system (using the existing computers and graphics capabilities), and the use of microprocessors in self-contained miniaturized modules that perform all functions involved in workload assessment.

Table 6 provides a summary of the seven critical criteria used to evaluate each of the various workload measurement approaches for the in-flight environment. This matrix provides some perspective on the relative feasibility of implementation, *provided the measurement technique could otherwise be perfected*. Details of these feasibility considerations are provided in the descriptions of each method which follow.

WORKLOAD TECHNIQUES SUMMARY

1. SUBJECTIVE OPINIONS

Subjective opinions are a commonly used measure of workload in flight test and evaluation. Often this measure is used in conjunction with other indices to provide a broader basis for evaluation and comparison. A variety of techniques exist for gathering subjective opinions. These include psychometrically defined rating scales, structured questionnaires with dichotomous or multiple choice responses, open-end questionnaries, structured interviews, and unstructured interviews.

In workload assessment applications, primarily two general approaches have been used. The more systematic approach deals with the use of rating scale procedures for obtaining pilot opinions; whereas, the second area deals with less structured approaches using a variety of interview and questionnaire procedures. Often the terms rating scale and questionnaire are used somewhat interchangeably in the scientific literature. For the purposes of this review, rating scales will be used for procedures which represent subject opinions gathered by devices with psychometric scaling properties, and questionnaires used in structured interviews will refer to procedures that are not based strictly on scaling considerations. Consequently, questionnaires have been grouped with interviews for the purposes of this review.

1.1 Rating Scales. Over the last twenty years much work has been dedicated to the development of rating scales for assessing the handling qualities of aircraft. These scales ordinarily contain about ten categories with descriptors that are not readily subject to confusion. The most widely used of these scales is the Cooper-Harper scale (1969). It is accepted for use in handling qualities work and is primarily used by test pilots. The descriptors of this scale pertain to the "flyability" of an aircraft. Even though the scale does contain some reference to workload the descriptors would have to be modified for use in workload applications. If this Cooper-Harper scale were used for workload assessment in its present form, the assumption must be made that handling difficulty and workload are directly related. Such an assumption may well be unwarranted.

Recently, some research has been directed toward the development and evaluation of workload-specific rating scales. Comparisons have been made between the workload measurements obtained from rating scales and those obtained from primary task performance, secondary tasks, occlusion, and physiological measures (Hicks and Wierwille, in press). Specifically, the rating scale proved to be a sensitive measure of workload and resulted in little intrusion on the primary task. Additional research has been directed toward developing a research-based, conjoint rating scale of workload for the F-18 aircraft (O'Conner and Buede, 1977; and Donnell and O'Conner, 1978) which was a direct outgrowth of the work of Helm (1975, 1976a, and 1976b).

With the exception of the conjoint measurement technique, most previous approaches have failed to follow rigorous psychometric procedures in developing workload rating scales. Examples of the use of ratings in this regard can be given both for flight simulator studies (e.g., Johannsen, 1976; Kreifeldt, Parkin, and Rothschild, 1976; Murphy and Gurman, 1972; and Schultz, Newell, and Whitbeck, 1970) and flight tests (e.g., Baker and Intano, 1974; Helm, 1975, 1976a; Lebacqz and Aiken, 1975; and Stackhouse, 1973).

1.2 Interviews and Questionnaires. In contrast to the rather rigorous procedures available for the development of rating scales, the procedures used in interviews and questionnaires are not nearly as structured. Application of these procedures to aircrew workload assessment range from completely open-ended debriefing sessions after flights (Soliday, 1965), to self-reporting logs of stressful activities (Soutendam, 1977; Cantell and Hartman, 1967), to carefully chosen questionnaire items (Steininger 1977). Recent work by Rohmert (1977) demonstrates procedures that can be employed in using questionnaire development. This approach, called the "Ergonimic Job Description Questionnaire," was developed specifically for workload evaluations of air traffic control activities.

If questionnaires and interviews are used in an unstructured or open-ended way, care still needs to be given to the appropriate topic areas and questions chosen for inclusion. If, on the other hand, structured responses are used, the choice of response items (e.g., dichotomous or multiple choice) should be constructed and tested much in the same manner as described for rating scales.

2. SPARE MENTAL CAPACITY

The largest body of research data dealing with the measurement of human operator workload is concerned with the evaluation of the concept of spare (residual or reserve) mental capacity. This concept is grounded on the fundamental assumption of a single-channel, sampling model of the human operator (Knowles, 1963; and Rolfe, 1973b). The approach assumes that an upper bound exists on the ability of the human operator to gather and process information. Spare mental capacity, then, is the difference between the total workload capacity of the operator and the capacity needed to perform the task. As spare mental capacity decreases, the operator's workload increases until a point of overload is reached. At this point, the information processing demands of the task exceed the operator's total workload capacity.

A variety of methods and procedures have been developed to measure, both directly and indirectly, spare mental capacity. In addition, a great deal of laboratory research data exist on empirical tests of various ramifications of the single-channel concepts. For example, data are available on the possibility of multi-channel processing; procedures for switching attention among channels; various points of conflict or bottlenecks in the human information processing channel; and variations in the upper limit of an individual's mental workload capacity due to factors of stress, emotional state, fatigue, and effort. Much of this human performance research is summarized by Kahneman (1973) and will not be reviewed at this time.

Essentially, three general methodological approaches have been advanced for measurement of workload using the generalized spare mental capacity paradigm. These approaches include task analytic, secondary task, and occlusion procedures. These methods are presented with the overall caution that even though the underlying single-channel, sampling model assumptions of the human operator is a viable concept, it is not a totally unequivocal hypothesis in terms of supporting data.

2.1 Task Analytic. Task analytic methods assess spare mental capacity by using mathematical/theoretical methods from systems engineering. The data base used in these techniques is most often obtained through laboratory and simulation experiments rather than flight tests. Task analytic methods assume that all task components, performed serially, require specified lengths of time to complete. As long as the actual time available for overall completion exceeds the sum of theoretical time durations for performing the task components, spare mental capacity exists. However, when the actual time available is insufficient, stress and task overloading occur. Task analytic methods consist of either task component/time summation computer models (Greening, 1978) or information-theoretic based procedures (Senders, 1970 and Baty, 1971).

2.2 Secondary Task. Most behavioral research approaches to estimating spare mental capacity have used secondary task procedures. This approach provides the human operator with an additional (secondary) task to be performed only when the main (primary) task has been fully attended. Performance on the secondary task theoretically decreases as the attentional demand of the primary task increases. Secondary task performance, then, becomes an indirect measure of operator workload.

Choice of the secondary task and procedures used to administer it become central issues in considering this method of workload assessment. Knowles (1963), for example, states that a viable secondary task for workload assessment should not physically interfere with the primary task, require little of scoring. Detailed reviews of the extensive literature on secondary tasks are provided by Rolfe (1973b) and Levine, Ogden and Eisner (1978).

2.3 Occlusion. In many cases where workload is to be estimated the primary information input to the operator is visual. The occlusion method of workload estimation can be used in such cases (Senders, et al., 1967).

Occlusion is a time-sharing technique and as such is similar to the secondary task method. However, in occlusion the time-sharing is accomplished by suppressing information inputs; that is, by giving the operator time samples of visual information. Examples of automobile driver research using this technique are found in Farber and Gallagher (1972) and Hicks and Wierwille (in press).

3. PRIMARY TASK MEASURES

It can be hypothesized that as the mental workload of a human operator increases, the performance of that operator may change, ordinarily in the direction of degradation. If such a change does in fact occur, its measurement would be an indication of increased workload. This hypothesis underlies the primary task performance method of assessing workload.

The use of primary task measures as a means of assessing workload was not particularly popular during the 1960's and early 1970's, because initial indications were that operators adapt to changing conditions, thereby holding performance constant. As Cooper and Harper (1969) put it, "In a specific task, he (the pilot) is capable of attaining essentially the same performance for a wide range of vehicle characteristics, at the expense of significant reductions in his capacity to assume other duties. . ." In this case they were referring to measures such as glide-slope error or flight path error in turbulence.

A somewhat more detailed examination of performance, however, might provide an indication of changes. As a task becomes more difficult, an operator may summon more effort, thereby holding performance in a specific variable or set of variables constant. However, to maintain this performance, the operator may have to modify his strategy. By examining measures other than those involving system output, it may be possible to detect this shift in strategy and thereby obtain a measure of workload.

Another concept in primary task measures was recently put forth by Albanese (1977). He suggests that "successful mission completion" is a measure of workload. In this case, if an operator is able to complete a mission successfully, there is no overload. On the other hand, if the operator cannot successfully complete the mission, an overload is presumed to have occurred. This rather broad concept has distinct merit if an investigator is most concerned about the overload/nonoverload dichotomy. Primary task measures properly chosen, will indeed make assessment of mission success possible. Measures such as landing touch-down performance, aiming performance, seeker lock-on, and number of procedural blunders, can be used. Successful mission completion must be defined in terms of the measures.

3.1 Single Measures (Primary Task). A very large number of workload studies (Murphy, et al. 1974; Price, 1975; and Wickens and Kassel, 1977) have involved the use of one or more primary task measures, individually on performance or as a precaution, while main interest was on some other method of assessing workload (Kalsbeek and Sykes, 1967 and Trumbo, et al. 1967). In a few cases, the primary measures have been taken specifically as a means of investigating level of workload (Brictson, 1974 a, b).

3.2 Multiple Measures (Primary Task). When a human operator performs a task in an actual system, several subtasks are ordinarily involved. In such cases, a single measure of system performance, such as error, may be inadequate. Considerations such as stores usage, accelerations experienced, and operator perceptual style and strategy may become important. In other situations, it may be found that single measures of the primary task do not exhibit adequate sensitivity to operator workload, because of operator adaptivity. In cases such as these, multiple measures of primary task variables might be considered for workload assessment. Essentially, the use of multiple measures provides a more complete picture of operator behavior and operator/system performance.

To obtain the maximum information, the multiple measures should first be subjected to a combined analysis and then subsequently to individual analysis where appropriate. Techniques that can be used for the combined analysis include multiple-regression analysis, correlation analysis, and various multivariate analyses. These techniques provide a sound methodological approach for drawing valid conclusions regarding system performance and workload.

Ordinarily, when using multiple measures, the additional measures used are not simply a greater number of those used in single measure analysis. Measures such as RMS accelerations, number of control (stick) reversals, dominant spectral frequencies, and control surface zero crossings are typical of the added measures (Kreifeldt, et al. 1976). Usually, measures such as these are intended to reflect strategy changes instead of performance scores, because performance scores may not change at lower operator workloads

In several cases multiple measures have been taken which combine several totally different workload assessment techniques. Primary task measures may be combined with any of the other methods: opinion, spare mental capacity, and physiological measures (Clement, 1976; O'Donnell and Spicuzza, 1975; and Simmons et al. 1976). The fact that the units of these measures differ does not present a problem in the analysis. The scores can be normalized or similarly treated in the analysis. In fact, a few studies have been performed with the purpose of determining which of several different workload techniques is most sensitive. (See Wierwille and Hicks, in press.)

3.3 Mathematical Modeling. Mathematical modeling of human operators in systems has long been an area of substantial interest to researchers. Interest began in the area of tracking and manual control system. Subsequently, it has branched into areas of human operator decision processes, supervisory processes, and team interactions.

Recently, a few of the researchers (Jex and Allen, 1970a; Baron and Levison, 1975; Wewerinke, 1976 and 1977; and Wickens and Gopher, 1977) involved in modeling have begun to examine the problem of operator workload. This has usually been done as an attendant examination, with prime interest being in model stimulus-response accuracy (Phatak, 1973; and Watson, 1972).

Other recent studies have departed from the describing function and optimal control models. Onstott and Faulkner (1977) (also Faulkner and Onstott, 1977) worked with an urgency model of attention allocation. Rouse (1977b) employed queuing theory to study human interaction with computers. Also Navon and Gopher (1977) postulate a model based on resource allocation. These models all have some bearing on workload; however, results are preliminary.

4. PHYSIOLOGICAL MEASURES

One of the most widely researched methods of assessing operator workload is the use of physiological measures. The physiological method generally involves the measurement and data processing of one or more variables related to human physiological processes. The underlying concept in physiological monitoring is as follows:

> As operator workload changes, involuntary changes take place in the physiological processes of the human body (body chemistry, nervous system activity, circulatory or respiratory activity, etc.). Consequently, workload may be assessed by the measurement and processing of the appropriate physiological variables.

In many cases, there is an underlying assumption that high workload levels are accompanied by increased emotional stress. This stress is then measured by physiological recording and is related back to workload. Stress in this case is assumed to act as an intermediate variable, causing physiological changes.

In other cases, the underlying assumption involves changes in the state of "arousal." Arousal may be considered as a state of preparedness of the body of level of activation of the human organism. Roughly, one may think of arousal as the state of excitedness. Here again, the assumption is that mental workload changes are accompanied by changes in arousal level that can be measured by appropriate physiological monitoring equipment.

It is worth mentioning that physiological measures of workload do not require the underlying assumption that the human operator is a single-channel sampling device. Instead, a rather global definition of workload may be assumed, in which mental workload is considered a conglomerate of behaviors, similar to those enumerated by Berliner, et al. (1964).

4.1 Single Physiological Measures. The majority of work on physiological monitoring for the sake of assessing workload has been performed using single measures. In several cases data on more than one measure have been taken in a given experiment, but each measure has then been analyzed individually. Such measures are considered here as single measures. Although perhaps not stated explicitly by the investigators, the objective of these studies has been to find a single physiological measure that accurately and reliably reflects changes in operator mental workload.

In dealing with single measures (or any physiological measures for that matter), it must be recognized that operator behavior other than mental workload may have an effect on the physiological measures. Physical exertion, for example, may affect the measures being taken. Consequently, the range of potential applications of a measure may be severely limited by the confounding effect of operator behavior in areas other than mental workload. In specific terms, a measure that varies with physical work as well as mental work for example can only be used if physical work is held constant or its manifestations on the measure are known and taken into account.

A review of each of the physiological measures as shown in Table 2 is beyond the scope of this paper. However, the reader is referred to a discussion of combined physiological measures in the following section 4.2 and Wierwille and Williges (1978).

4.2 Combined Physiological Measures. Certain investigators have taken the point of view that single physiological measures may not provide adequate predictive information to allow assessment of workload. They then proceed to analyze multiple physiological measures in a combined analysis in an effort to better assess and predict workload. The multiple physiological measurement philosophy is the same approach taken by researchers as was discussed in Section 3.2 for multiple primary task measures.

As with primary task measures, a common class of techniques can be applied. These include multiple-regression analysis, correlation analysis, and multivariate analysis. The purpose in using these statistical techniques is to provide the best prediction and discrimination of workload levels, based on the physiological measures at hand.

Several reports and papers have been published describing multiple feature extraction techniques applied to multiple physiological measures (Spyker, et al. 1971; Stachouse, 1973; and Stackhouse, 1978). The technique used is one of selecting a number of features for each physiological measure and then performing a multiple regression. The best weighting of the most highly correlated features is then used in the prediction equation.

Storm, et al. (1976) have performed analysis of multiple compounds in the urine believed correlated with various aspects of aircrew-member stress. In general, while statistical analyses were not performed, great care was taken in analyzing from a diagnostic point of view the directions and magnitudes of changes in the levels of the compounds. Moreover, interactions were studied. A study of this type as well as Storm and Hapenney (1976) gives a general impression of the physiological changes that occur when Air Force aviators undergo high workload/stress conditions for extended periods of time. Brictson, et al. (1974) and McHugh, et al., (1974) studied the effects of high workload conditions on the performance of naval aviators in high-performance aircraft. The approach taken was one which combined stepwise multiple regression of physiological, psychiatric, and performance measures in carrier landings.

The physiological measures in these studies on naval aviators were primarily those taken from blood samples and included serum cholestrol, serum uric acid, blood lactate, and pyruvate. Changes in the levels of biochemical measures were analyzed as a function of alterations in levels of workload, sleep, performance, and mood.

4.3 <u>Speech Pattern Analysis</u>. Recently, there have been indications that inaudible changes take place in speech when an individual is under stress. These changes generally are not detectable by an unaided listener but can be elicited with the proper equipment, e.g., Psychological Stress Evaluator (PSE). The underlying theory of the PSE has to do with presence or absence of physiological tremor of micro-tremor in the human voice. In general this micro-tremor is present in an individual who is <u>not</u> under stress. The tremor results in a frequency modulation effect of certain voice sounds that is only detectable with equipment. The tremor and frequency modulation of the voice become suppressed when an individual is under stress, such as when attempting to deceive law enforcement personnel (Kradz, 1974, and Dahm, 1974).

Older and Jenney (1975) analyzed voice communications of Skylab astronauts as a means of determining situational stress. The scores obtained using a PSE were correlated with operational variables known to represent varying degrees of stress. They found some statistically significant relationships, but concluded that PSE usage was not sufficiently predictive of mild stress as to warrant use in future missions.

Simonov and Frolov (1977), following the work of Older and Jenney, undertook to determine the emotional state of cosmonauts and others via voice analysis. They indicated that the problem appears very complex and that substantial further work is required.

Harris, et al (1977), taking a somewhat different approach, using automatic voice recognition and synthesis equipment, showed that a verbal arithmetic task produced less decrement in concurrent manual tracking than did a keyboard arithmetic task. They point out that automatic voice recognition equipment introduces an additional source of error that may be dependent on task difficulty.

It seems clear that extreme stress can be measured by voice analysis. At this time, however, the usefulness of voice analysis for either mild stress or mental workload is unclear. Several investigators appear on the verge of analysis of voice in regard to workload, but results are not presently available.

CONCLUSIONS

This survey of the workload literature has shown that several approaches are potentially useful for the aircrew workload problem, but no one single technique can be recommended as <u>the</u> definitive measure of operator workload. Because of the multidimensionality of workload, it also appears unlikely that any one single measure will ever suffice completely. Consequently, multiple measures including the dimensions of subjective opinions, spare mental capacity, primary tasks, and physiological correlates should be considered. The classification scheme and applicability matrix developed in this paper should provide the investigator with an aid for choosing among the presently available techniques.

RECOMMENDATIONS

This study of the workload literature has provided support for several recommendations, including implications for future work. Four of the most prominent research recommendations are presented in brief [illegible].

<u>Computerized Information Retrieval System for Workload Literature</u>

This study of the workload literature has been performed in a way that will allow computerizing of the information. The advantages of computerizing would be numerous. A user would be guided through important citations based on the needs associated with a given aircrew workload estimation problem. More specifically, relevant references could be cross-filed according to:

1. The workload classification scheme.
2. Keyword or combinations of keywords (in title or in abstract).
3. Author or authors.
4. Workload category or subcategory.

If requested by the user, the system would also provide a narrative summary on the N workload techniques. to provide broad necessary background should the user not already have it.

<u>Improvement of Rating Scales for Workload Estimation</u>

The two major advantages of subjective opinion ratings are acceptance and lack of intrusion. Pilot acceptance of opinion ratings has been good and is well documented in the handling quality domain. Opinion ratings are generally not intrusive. However, with the exception of the conjoint measurement technique, most previous approaches have failed to follow rigorous psychometric procedures in developing workload rating scales. Additionally, several other limitations of ratings also need to be considered.

Adaptivity of the pilot, for example, represents a serious problem. Due to adaptivity, ratings may be either too high or too low. A system that initially provides the impression of being awkward to use may obtain higher ratings than it should, because the crew member adapts. Other problems include possible emotional state, experience, and learning.

Given the widespread use and general applicability of rating scales as a technique of workload assessment, it is surprising that a rigorous workload rating scale has not been developed. Research is needed to determine the underlying scaling dimensions of mental workload and to develop an interval type metric characteristic of the conjoint measurement procedure. Recent approaches such as behaviorally anchored response scales (BARS) may be useful in this regard. Objective anchor points such as semantic differential such as policy capturing might be applicable in determining the relative importance of various dimensions used in subjective estimates of workload. Research is also needed to compare the utility of these various rating procedures and to specify the reliability and validity of the resulting scales.

Comparison of Methods of Workload Estimation

This literature review has shown that little work has been done on experimental comparison of workload estimation methods. To a great extent each research group in the workload estimation area tends to advocate usually one or possibly two workload estimation techniques. One group advocates time-estimation, another critical tracking tasks, and still others specific kinds of physiological measures. While all this work is clearly important, particularly in regard to development, evaluation, and optimization of various techniques unanswered.

Hicks and Wierwille (in press) have recently addressed this problem on an initial basis. They compared five different (specific) workload techniques in a moving-base driving simulator. Included in their comparison were rating scales, primary task measures, secondary task measures, occlusion, and heart rate variability. It was found that large differences in technique sensitivity existed when operator loading was adjusted under controlled conditions. Sensitivity in this context is defined as the statistically significant differences in operator loading. High sensitivity low variance of the scores about the means. In addition, it was determined that the degree of intrusion varied with the technique, with some being nonintrusive while others were highly intrusive.

A similar, more complete study needs to be performed for the aircrew workload estimation problem. At present, the comparative sensitivity of aircrew workload estimation techniques is unknown. Because sensitivity has generally not been high, such a study is vital. Selection of a technique without comparative information may yield results indicating that there is no change in aircrew workload for two or more different configurations when in fact there is a change. And, since an aircrew member's workload may already be high, failure to discriminate workload differences in a T&E situation may later jeopardize mission success.

Update of Literature Search on Workload Evaluation Techniques

Workload evaluation is at present a highly active research area. It is estimated that more than one hundred researchers in the United States, Europe, and elsewhere are immersed in workload research at this time. Because of the forthcoming results and the extreme diversity of this work, the workload search described here will need to be updated peridocially if it is to remain current.

The updating of the search is very important since much of the work presently in progress has direct bearing on the aircrew workload problem. More specifically, while much of the earlier research on workload was of an exploratory nature or involved development of concepts and constructs, more recent work has tended toward the practical with applications to aircraft and other human-operator systems problems.

APPENDIX

WORKLOAD BIBLIOGRAPHY

Albanese, R. A. Mathematical analysis and computer simulation in military mission workload assessment. Proceedings of the AGARD Conference on Methods to Assess Workload, AGARD-CPP-216, April, 1977, A13-1 - A13-6.

Allen, R. W., Jex, H. R., McRuer, D. T. and DiMarco, R. J. Alcohol effects on driving behavior and performance in a car simulator. IEEE Transactions on Systems, Man, and Cybernetics, 1975, SMC-5, 498-505.

Allport, D. A., Antonis, B. and Reynolds, P. On the division of attention: a disproof of the single channel hypothesis. Quarterly Journal of Experimental Psychology, 1972, 24, 225-235.

Alluisi, E. A. and Morgan, B. B., Jr. Effects of practice and work load on the performance of a code transformation task (COTRAN). Moffett Field, California: National Aeronautics and Space Administration, Contractor's Report NASA CR-1261, 1969.

Alluisi, E. A. and Morgan, B. B., Jr. Effects of sustained performance of time-sharing a three-phase code transformation task (3P-COTRAN). Perceptual and Motor Skills, 1971, 33, 639-651.

Anderson, P. A. and Toivanen, M. L. Effects of varying levels of autopilot assistance and workload on pilot performance in the helicopter formation flight mode. Washington, D.C.: US Office of Naval Research, JANAIR 680610, March, 1970.

Armstrong, G. C., Sams, D. D., McDowell, J. W. and Winter, F. J., Jr. Pilot factors for helicopter pre-experimental phase. Randolph AFB, Texas: USAF Instrument Flight Center, IFC-TR-74-2, February, 1975.

Asiala, C. F. Advanced man-machine evaluation techniques. Paper presented at the American Defense Preparedness Association, Huntsville, Alabama, November 12-13, 1975.

Asiala, C. F., Loy, S. L. and Quinn, T. J. Digital simulation model for fighter pilot workload. St. Louis, Missouri: McDonnell Aircraft Company, MDC A0058, September, 1969.

Auffret, R., Seris, H., Berthoz, A. and Fatras, B. Estimate of the perceptive load by variability of rate of heartbeat: Application to a piloting task. Le Travail Humain, 1967, 30, 309-310.

Bahrick, H. P., Noble, M. and Fitts, P. M. Extra-task performance as a measure of learning a primary task. Journal of Experimental Psychology, 1954, 48, 298-302.

Bainbridge, L. Forgotten alternatives in skill and work-load. Ergonomics, 1978, 21, 169-185.

Baker, D. L., and Intano, G. P. Helicopter yaw axis augmentation investigation - CDG-PFH-4. Randolph AFB, Texas: USAF Instrument Flight Center, IFC Test Plan 74-11, December 1974.

Barnes, J. A. Use of eye-movement measures to establish design parameters for helicopter instrument panels. Proceedings of the AGARD Conference on Methods to Assess Workload, AGARD-CPP-216, April 1977, A3-1-A3-8.

Baron, S., and Levison, W. H. An optimal control methodology for analyzing the effects of display parameters on performance and workload in manual flight control. IEEE Transactions on Systems, Man, and Cybernetics, 1975, SMC-5, 423-430.

Bate, A. J., and Self, H. C. Effects of simulated task loading on side-looking radar target recognition. Wright-Patterson AFB, Ohio: Aerospace Medical Research Laboratory, AMRL-TR-67-141, June 1968.

Baty, D. L. Human transinformation rates during one-to-four axis tracking with a concurrent audio task. Proceedings of the 7th Annual NASA-University Conference on Manual Control, University of Southern California, June 1971, 293-306 (NASA SP-281).

Beatty, J. Pupillometric measurement of cognitive workload. Proceedings of the 12th Annual NASA-University Conference on Manual Control, University of Illinois, May 1976, 135-143 (NASA TM X-73 70).

Benson, A. J., Huddleston, J. H. F., and Rolfe, J. M. A psychophysiological study of compensatory tracking on a digital display. Human Factors, 1965, 7, 457-472.

Bergeron, H. P. Pilot response in combined control tasks. Human Factors, 1968, 10, 277-282.

Bergstroem, B., and Arnberg, P. Heart rate and performance in manual missile guidance. Perceptual and Motor Skills, 1971, 32, 352-354.

Beyer, R. A study of pilot's workload in helicopter operation under simulated IMC employing a forward looking sensor. Proceedings of AGARD Conference on Studies on Pilot Workload, AGARD-CPP-217, April 1977, B6-1-B6-10.

Bisseret, A. Analysis of mental processes involved in air traffic control. Ergonomics, 1971, 14, 565-570.

Borg, G. Subjective aspects of physical and mental load. Ergonomics, 1978, 21, 215-220.

Boyce, P. R. Sinus arrhythmia as a measure of mental load. Ergonomics, 1974, 17, 177-183.

Boylan, R. J. A review of crew systems analytic methods. Seattle, Washington: Boeing Aerospace Company, D160-17525-1, January 1974 (a).

Boylan, R. J. Introduction to Boeing operator workload and workspace evaluation models. Seattle, Washington: Boeing Aerospace Company, D180-17526-1, January, 1974. (b)

Bradshaw, J. L. Load and pupillary changes in continuous processing tasks. British Journal of Psychology, 1968, 59, 265-271.

Brichcin, M. and Hampejsova, O. Results of two kinds of mental load measurements. Ceskoslovenska Psychologie, 1970, 14, 19-31. (In Czechoslovakian).

Bricteon, C. A. Pilot landing performance under high workload conditions. Proceedings of the AGARD Conference on Simulation and Study of High Workload Operations, AGARD-CP-146, April, 1974, A7-1 - A7-10. (a)

Bricteon, C. A. Pilot landing performance under high workload conditions. La Jolla, California: Dunlap and Associates, Contract N00014-73-C-0053, April, 1974. (AD/A 001 802). (b)

Bricteon, C. A. Methods to assess pilot workload and other temporal indicators of pilot perforance effectiveness. Proceedings of AGARD Conference on Studies on Pilot Workload, AGARD-CPP-217, April, 1974, B10-1 - B10-7.

Bricteon, C. A., McHugh, W. and Naitoh, P. Prediction of pilot performance: Biochemical and sleep-mood correlates under high workload conditions. Proceedings of the AGARD Conference on Simulation and Study of High Workload Operations, AGARD CP-146, April, 1974, A13-1 - A13-10.

Brigham, F. R. COPTEC-A controller overload prediction technique. Loughborough, England: Human Sciences and Advanced Technology Research Group, Department of Human Sciences, University of Technology, HUSAT Memo. No. 67, February, 1974.

Broadbent, D. E. and Heron, A. Effects of a subsidiary task on performance involving immediate memory by younger and older men. British Journal of Psychology, 1962, 53, 189-198.

Bromberger, R. A. LAMPS simulations: VI Pilot performance in the LAMPS simulator. Warminster, Pennsylvania: Naval Air Development Center, NADC-76191-40, October, 1976.

Brown, E. L., Stone, G. and Pearce, W. E. Improving cockpits through flight crew workload measurement. Paper presented at the 2nd Advanced Aircrew Display Symposium, U.S. Naval Air Test Center, Patuxent River, Maryland, April 23-25, 1975. (Douglas Paper 6355).

Brown, I. D. Measuring the spare mental capacity of car drivers by a subsidiary auditory task. Ergonomics, 1962, 5, 247-250.

Brown, I. D. A comparison of two subsidiary tasks used to measure fatigue in car drivers. Ergonomics, 1965, 8, 467-473.

Brown, I. D. Subjective and objective comparisons of successful and unsuccessful trainee drivers. Ergonomics, 1966, 9, 49-56.

Brown, I. D. Dual task methods of assessing workload. Ergonomics, 1978, 21, 221-224.

Brown, I. D. and Poulton, E. C. Measuring the spare mental capacity of car drivers by a subsidiary task. Ergonomics, 1961, 4, 35-40.

Burke, J. E. Use of Eye Mark/Sony Videocorder System and related data reduction. Dallas, Texas: LTV Aerospace, Vought Systems Division, VSD Report 2-57110/3R-3107, August, 1973.

Burke, J. E. In flight acquistion of task sequences and task times. Paper presented at the meeting of the Aerospace Medical Association, Las Vegas, Nevada, May, 1977.

Cannings, R., Borland, R. G., Hill, L. E. and Nicholson, A. N. Pitch and formant analysis of the voice in the investigation of pilot workload. Proceedings of the AGARD Conference on Methods to Assess Workload, AGARD-CPP-216, April, 1977, A5-1 - A5-10.

Cantell, G. K. and Hartman, B. O. Application of time and workload analysis technics to transport flyers. Brooks AFB, Texas: USAF School of Aviation Medicine, Technical Report SAM-TR-67-71, August, 1967.

Caplan, R. D. and Jones, K. W. Effects of workload, role ambiguity, and type a personality on anxiety, depression, and heart rate. Journal of Applied Psychology, 1975, 60, 713-719.

Casey, S. M., Brietmaier, W. A. and Nason, W. E. Cerebal activation and the placement of visual displays. Warminster, Pennsylvania: U.S. Naval Air Development Center, NADC-77247-40, August, 1977.

Catlett, R. L. Application of information theory concepts to study work complex versus operator action time. Red River Army Depot, Texarkana, Texas: USAMC Intern Training Center, USAMC-ITC-2-73-08, March, 1973. (AD 786 286).

Ceder, A. Driver's eye movements as related to attention in simulated traffic flow conditions. *Human Factors*, 1977, *19*, 571-581.

Chainova, L. D., Komarova, I. A. and Zonabend, F. I. Complex psychophysiological evaluation of the readability of symbolic information (Kompleksnaya Psikhofiziologichesdaya Ot senka Chitaemosti Znakovoy Informatsii). *Voprosy Psikhologii*, 19[illegible], *16*, 163-168. (Royal Aircraft Establishment Library Translation 1777).

Chiles, W. D. Objective methods for developing indices of pilot workload. Oklahoma City, Oklahoma: Federal Aviation Administration, Civil Aeromedical Institute, FAA-AM-77-15, July, 1977.

Chiles, W. D. Objective methods. In A. H. Roscoe (Ed.) *Assessing pilot workload*. AGARD-AG-233, February, 1978.

Chiles, W. D. and Alluisi, E. A. A review of methods for specifying operator or occupational workload. Paper presented at XIXth International Congress of Applied Psychology, Munich, Germany, August, 1978.

Chu, Y. and Rouse, W. B. Optimal adaptive allocation of decision making responsiblity between human and computer in multi-task situations. *Proceedings of the 1977 International Conference on Cybernetics and Society*, Washington, D.C., September, 1977.

Clark, W. E., Jr. and Armstrong, G. C. Three-cue helicopter flight director evaluation. Randolph AFB, Texas: USAF Instrument Flight Center, IFC TR 77-3, July, 1977.

Clement, W. F. Investigating the use of a moving map display and a horizontal situation indicator in stimulated powered-lift short-haul operations. *Proceedings of the 12th Annual NASA-University Conference* (NASA-TMX-73 70).

Clement, W. Annotated bibliography of procedures which assess primary task performance in some manner as the basic element of a workload Technical Report No. 1104-2, January, 1978.

Clement, W. F., Jex, H. R. and Graham, D. A manual control-display theory applied to instrument landings of a jet transport. *IEEE Transactions on Man-Machine Systems*, 1968, *MMS-9*, 93-109. (a)

Clement, W. F., Jex, H. R. and Graham, D. Application of a systems analysis theory for manual control displays to aircraft instrument landing. *Proceedings of the 4th Annual NASA-University Conference on Manual Control*, Ann Arbor, Michigan, March, 1968, 69-94. (NASA SP-192). (b)

Clement, W. F., McRuer, D. T. and Klein, R. H. Systematic manual control display design. *Proceedings of the AGARD Conference on Guidance and Control Displays*, AGARD-CP-96, 6-1 - 6-10, 1972.

Cliff, R. C. The effects of attention sharing in a dynamic dual-task environment. *Proceedings of the 7th Annual NASA-University Conference on Manual Control*. University of Southern California, June, 1971. (NASA SP-281).

Cohen, S. I. and Silverman, A. J. Measurement of pilot mental effort. Paris, France: North Atlantic Treaty Organization, Advisory Group for Aeronautical Research and Development, Report 148, May, 1957.

Colle, H. A. and DeMaio, J. C. The use of dual-task performance operating curves to assess workload. Paper presented at the 1978 Review of Air Force Sponsored Basic Research in Flight and Technical Training, U.S. Air Force Academy, Colorado Springs, Colorado, April, 1978.

Control-display pilot factors program. Randolph AFB, Texas: USAF Instrument Pilot Instructor School, Instrument Evaluation Project NR 63-1, December, 1963.

Cooper, G. E. and Harper, R. P., Jr. The use of pilot rating in the evaluation of aircraft handling qualities. Moffett Field, California: National Aeronautics and Space Administration, Ames Research Center, NASA TN-D-5153, April, 1969.

Corkindale, M.G.G. A flight simultor study of missile control performance as a function of concurrent workload. *Proceedings of the AGARD Conference on Simulation and Study of High Workload Operations*, AGARD-CP-146, April, 1974, A5-1 - A5-6.

Corlett, E. N. Cardiac arrhythmia as a field technique: Some comments on a recent symposium. *Ergonomics*, 1973, *16*, 3-4.

Couluris, G. J., Ratner, R. S., Petracek, S. J., Wong, P. J. and Ketchel, J. M. Capacity and Productivity implications on en route air traffic control automation. Washington, D.C.: Federal Aviation Administration, FAA-RD-74-196, December, 1974. (AD/A 016 622).

Crabtree, M. S. Human factors evaluation of several control system Configurations, including workload sharing with force wheel steering during approach and flare. Wright-Patterson AFB, Ohio: USAF Flight Dynamics Laboratory, AFFDL-TR-75-43, April, 1975.

Crawford, B. M., Pearson, W. H. and Hoffman, M. Multifunction switching and flight control workload. *Paper presented at the 6th Psychology in the DoD Symposium*, U.S. Air Force Academy, Colorado Springs, Colorado, April, 1978.

Faulkner, W. H. and Onstott, E. D. Error rate information in attention allocation pilot models. Proceedings of the 13th NASA-University Conference on Manual Control, Massachusetts Institute of Technology, June 15-17, 1977, 72-78.

Finkelman, J. M. and Glass, D. C. Reappraisal of the relationship between noise and human performance by means of a subsidiary task measure. Journal of Applied Psychology, 1970, 54, 211-213.

Firth, P. A. Psychological factors influencing the relationship between cardiac arrhythmia and mental load. Ergonomics, 1973, 16, 5-16.

Flora, C. C., Kriechbaum, G.K.L. and [illegible], W. A flight investigation of systems developed for reducing pilot workload and improving tracking accuracy during noise-abatement landing approaches. Moffett Field, California: National Aeronautics and Space Administration, Ames Research Center, Contractor's Report NASA CR-1427, 1969.

Fowler, R. L., Williams, W. E., Fowler, M. G. and Young, D. D. An investigation of the relationship between operator performance and operator panel layout for continuous tasks. Wright-Patterson AFB, Ohio: Aerospace Medical Research Laboratory, AMRL-TR-68-170, December, 1968.

Gabriel, R. F., and Burrows, A. A. Improving time-sharing performance of pilots through training. Human Factors, 1968, 10, 33-40.

Gardner, R. M., Beltramo, J. S., and Krinsky, R. Pupillary changes during encoding, storage, and retrieval of information. Perceptual and Motor Skills, 1975, 41, 951-955.

Gartner, W. B. and Murphy, M. R. Pilot workload and fatigue: A critical survey of concepts and assessment techniques. Moffett Field, California: National Aeronautical and Space Administration Ames Research Center, NASA TN D-8365, November, 1976.

Gaume, J. G. and White, R. T. Mental workload assessment, II. Physiological correlates of mental workload: Report of three preliminary laboratory tests. St. Louis, Missouri: McDonnell Douglas Corporation, Report MDC J7023/01, December, 1975.

Geer, C. W. Navy manager's guide for the test and evaluation sections of MIL-H-46855. Seattle Washington: Boeing Aerospace Company, Technical Report D194-10006-1, June, 1977. (b)

Geiselhart, R., Schiffler, R. J. and Ivey, L. J. A study of task loading using a three-man crew on a KC-135 aircraft. Wright-Patterson AFB, Ohio: Aeronautical Systems Division, ASD-TR-76-19, October, 1976.

Geiselhart, R., Koeteeuw, R. I. and Schiffler, R. J. A study of task loading using a four-man crew on a KC-135 aricraft (GIANT BOOM). Wright-Patterson AFB, Ohio: Aeronautical Systems Division, ASD-TR-76-33, April, 1977.

Gerathewohl, S. J. Definition and measurement of perceptual and mental workload in aircrews and operators of Air Force weapon systems: A status report. In Higher mental functioning in operational environments. North Atlantic Treaty Organization, April, 1976. (AD-A025 663).

Gerathewohl, S. J., Brown, E. L., Burke, J. E., Kimball, K. A., Lowe, W. F., and Stackhouse, S. P. Inflight measurement of pilot workload: A panel discussion. Aviation, Space, and Environmental Medicine, June, 1978, 810-822.

Glenn, F. A., Streib, M. I., and Wherry, R. J., Jr. The human operator simulator volume VIII: Applications to assessment of operator loading, Willow Grove, Pennsylvania: Analytics, Technical Report 1233-A, June, 1977.

Goerres, H.P. Subjective stress assessment as a criterion for measuring the psychophysical workload on pilots. Proceedings of the AGARD Conference on Studies on Pilot Workload, AGARD-CPP-217, April, 1977, B12-1 - B12-8.

Gopher, D. Eye movement patterns in selective listening tasks of focused attention. Perception and Psychophysics, 1973, 14, 259-264.

Gopher, D., Navon, D., Chillag, N. and Dotan, H. Tracking in two dimensions as a function of dimension priorities and tracking difficulty. Haifa, Israel: Technion-Israel Institute of Technology, The Center for Industrial Safety Research, Technical Report AFOSR-77-2, December, 1977.

Gopher, D. and North, R. A. Manipulating the conditions of training in time-sharing performance. Human Factors, 1977, 19, 583-593.

Graham, D. K. Transport airplane flight deck development survey and analysis: Report and recommendations. Moffett Field, California: National Aeronautical and Space Administration, NASA CR-145, 121, January, 1977.

Green, R. and Flux, R. Auditory communication and workload. Proceedings of the AGARD Conference on Methods to Assess Workload, AGARD-CPP-216, April, 1977, A4-1 - A4-8.

Greening, C. P. Analysis of crew/cockpit models for advanced aircraft. China Lake, California: Naval Weapons Center, NWC TP 6020, February, 1978.

Gregoire, H. G. Is man the weakest link? Proceedings of AGARD Conference on Methods to Assess Workload, AGARD-CPP-216, April, 1977, A1-1 - A1-3.

Gutmann, H. E., Easterling, R. G. and Webster, R. G. The effects of flicker on performance as a function of task-loading. Albuquerque, New Mexico: Sandia Laboratories, SC-TM-72 0617, November, 1972.

Hacker, W. Determining the psychic workload: Present status and perspectives. Socialistische Arbeitswissenschaft, 1974, 18, 17-28. (In German).

Hacker, W., Plath, H. E., Richter, P. and Zimmer, K. Internal representation of task structure and mental load of work: approaches and methods of assessment. Ergonomics, 1978, 187-194.

Hale, H. B., Anderson, C. A., Williams, E. W. and Tanne, E. Endocrine-metabolic effects of unusually long or frequent flying missions in C-130E or C-135B aircraft. Aerospace Medicine, 1968, 39, 561-570.

Hale, H. B., Hartman, B. O., Harris, D. A., Williams, E. W., Miranda, R. E., Hosenfeld, J. M. and Smith, B. N. Physiologic stress during 50-hour double-crew missions in C-141 aircraft. Brooks AFB, Texas: USAF School of Aerospace Medicine, SAM-TR-71-487, October, 1971.

Hale, H. B., Hartman, B. O., Harris, D. A., Williams, E. W., Miranda, R. E., Hosenfeld, J. M. and Smith, B. N. Physiologic stress during 50-hour double-crew missions in C-141 aircraft. Aerospace Medicine, 1972, 43, 293-299.

Hale, H. B., Hartman, B. O., Harris, D. A., Williams, E. W., Miranda, R. E. and Hosenfeld, J. M. Time zone entrainment and flight stressors as interactants. Aerospace Medicine, 1972, 43, 1089-1094.

Hale, H. B., McNee, R. C., Ellis, J. P., Jr., Bollinger, R. R. and Hartman, B. O. Endocrine-metabolic indices of aircrew workload: An analysis across studies. Brooks AFB, Texas: USAF School of Aerospace Medicine, Unpublished report.

Hale, H. B., McNee, R. C., Ellis, J. P., Jr., Bollinger, R. R. and Hartman, B. O. Endocrine-metabolic indices of aircrew workload: An analysis across studies. Proceedings of the AGARD Conference on Simulation and Study of High Workload Operations, AGARD-CP-146, April, 1974, A10-1 - A10-6.

Hale, H. B., Williams, E. W., Smith, B. N., and Melton, C. E., Jr. Excretion patterns of air traffic controllers. Aerospace Medicine, 1971, 42, 127-138.

Hall, T. J., Passey, G. E. and Meighan, T. W. Performance of vigilance and monitoring tasks as a function of workload. Wright-Patterson AFB, Ohio: Aerospace Medical Research Laboratories, AMRL-TR-65-22, March, 1965.

Hamilton, P. Process entropy and cognitive control: mental load in internalized thought processes. Position paper prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977.

Harris, D. A., Pegram, G. V. and Hartman, B. O. Performance and fatigue in experimental double-crew transport missions. Aerospace Medicine, 1971, 42, 980-986.

Harris, S. D., North, R. A. and Owens, J. M. A system for the assessment of human performance in concurrent verbal and manual control tasks. Paper presented at the 7th Annual Meeting of the National Conference on the Use of On-Line Computers in Psychology, Washington, D.C., November 9, 1977.

Hart, S. G. A cognitive model of time perception. Paper presented at the 56th Annual Meeting of the Western Psychological Association, Los Angeles, California, April, 1976.

Hart, S. G. and McPherson, D. Airline pilot time estimation during concurrent activity including simulated flight. Paper presented at the 47th Annual Meeting of the Aerospace Medical Association, Bal Harbour, Florida, May, 1976.

Hart, S. G., McPherson, D., Kriefeldt, J. and Wempe, T. E. Multiple curved descending approaches and the air traffic control problem. Moffett Field, California: National Aeronautical and Space Administration, Ames Research Center, NASA TM-78, 430, August, 1977.

Hart, S. G. and Simpson, C. A. Effects of linguistic redundancy on synthesized cockpit warning message comprehension and concurrent time estimation. Proceedings of the 12th Annual NASA-University Conference on Manual Control, University of Illinois, May, 1976. (NASA TM X-73 70).

Hartman, B. O., Hale, H. B. and Johnson, W. A. Fatigue in FB-111 crew-members. Aerospace Medicine, 1974, 45, 1026-1029.

Helander, M. G. Physiological reactions of drivers as indicators of road traffic demand. In Driver performance studies: Transportation Research Record 530. Washington, D.C.: U.S. Transportation Research Board, Technical Report TRB/TRR-530, 1975, 1-17.

Helm, W. R. Human factors test and evaluation, functional description inventory as a test and evaluation tool development and initial validation study. Volume I and II. Patuxent River, Maryland: U.S. Naval Air Test Center, SY-77R-75, September, 1975.

Helm, W. R. Human factors evaluation of model P-3C UPDATE I airplane: Third interim report. Patuxent River, Maryland: U.S. Naval Air Test Center, SY-122R-75, February, 1976(a).

Helm, W. R. Function description inventory as a human factors test and evaluation tool: An empirical validation study. Patuxent River, Maryland: U.S. Naval Air Test Center, SY-127R-76, July, 1976(b).

Helm, W. R. The application of computer aided evaluative techniques to system test and evaluation. Proceedings of the 21st Annual Meeting of the Human Factors Society, San Francisco, California, October, 1978, 92-94.

41

Cross, K. D. and Cavallero, F. R. Utility of the vertical contact analog display for carrier landings - a dignostic evaluation. *Proceedings of the AGARD Conference on Guidance and Control Displays*, AGARD-CP-96, 21-1 - 21-11, 1972.

Curry, R. E. Position paper prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977.

Dahm, A. E. Study of the field use of the psychological stress evaluator. Dektor Counterintelligence and Security, Inc., Springfield, Virgina, Unpublished manuscript, 1974.

Damos, D. and Wickens, C. A quasi-linear control theory analysis of time-sharing skills. *Proceedings of the 13th Annual NASA-University Conference on Manual Control*, Massachusetts Institute of Techonology, June 15-17, 1977. (a)

Damos, D. L. and Wickens, C. D. Dual-task performance and the Hick-Hyman law of choice reaction time. *Journal of Motor Behavior*, 1977, *9*, 209-215. (b)

Danev, S. G. and Wartna, G. F. Information load and time stress: Some psychophysiological consequences. *TNO-Nieuws*, 1970, *25* 389-395.

Daniel, J. Newer approaches to a research of mental load. *Proceedings of the 2nd meeting of Psychologists from the Danubian Countries*, Smolenice, Czechoslovakia, September, 1970.

Daniels, A. F. Crew workload sharing assessment in all-wheather, low-level strike aircraft. In *Problems of the Cockpit Environment*, AGARD Conference Proceedings No. 55, March, 1970. (AD 705 369).

Defayolle, M., Dinand, J. P., and Gentil, M. T. Averaged evoked potentials in relation to attitude, mental load, and intelligence. In W. T. Singleton, J. G. Fox, and D. Whitfield (Eds.) *Measurement of man at work*. London: Taylor and Francis, 1973, 81-91.

Dick, A. O. and Bailey, G. A comparison between oculometer data and pilot opinion on the usefulness of instruments during landing. Rochester, New York: University of Rochester, Center for Visual Science, Technical Report No. 3-76, 1976.

Dick, A. O., Brown, J. L. and Bailey, G. Statistical evaluation of control inputs and eye movements in the use of instrument clusters during aircraft landing. Rochester, New York: University of Rochester, Center for Visual Science, Technical Report 4-76, 1976.

Donnell, M. L. and O'Connor, M. F. The application of decision analytic techniques to the test and evaluation phase of the acquisition of a major air system: Phase II. McLean, Virginia: Decisions and Designs, Technical Report TR 78-3-25, April, 1978.

Dougherty, D. J., Emery, J. H. and Curtin, J. G. Comparison of perceptual workload in flying standard instrumentation and the contact analog vertical display. Washington, D.C.: Joint Army Navy Aircraft Instrumentation Research, D228-421-019, December, 1964.

Drennen, T. G., Curtin, J. G. and Warner, H. D. Manual control in target tracking tasks as a function of control type, task loading, and vibration. St. Louis, Missouri: McDonnell Douglas Corporation, MDCE 1713, August, 1977.

Dunn, R. S., Gilson, R. D. and Sun, P. A simulator study of helicopter pilot workload reduction using a tactile display. *Proceedings of the 12th Annual NASA-University Conference on Manual Control*, University of Illinois, May, 1976. (NASA TMX-73 70).

Dyer, R. F., Matthews, J. J., Wright, C. E. and Yudawitch, K. L. Questionnaire construction manual. Fort Hood, Texas: U.S. Army Research Institute for the Behavioral and Social Sciences, Field Unit, Technical Report P-77-1, July, 1976.

Edson, R. K. The Dektor psychological stress evaluator (voice stress analyzer) as a research instrument. Unpublished master's thesis, National Graduate University, April, 1976.

Ellis, G. A. Subjective assessment. In A. H. Roscoe (Ed.) *Assessing pilot workload*. AGARD-AG-233, February, 1978.

Enstrom, K. D. and Rouse, W. B. Telling a computer how a human has allocated his attention between control and monitoring tasks. *Proceedings of the 12th Annual NASA-University Conference on Manual Control*, University of Illinois, May, 1976, 104-123. (NASA TM X-73, 70).

Ephrath, A. R. *Pilot performance in zero-visibility precision approach* (NASA CR-137759). Doctoral dissertation, Massachusetts Insitute of Technology, 1975.

Ephrath, A. R. A novel approach to the cross-adaptive auxiliary task. *Proceedings of the 12th Annual NASA-University Conference on Manual Control*, University of Illinois, May, 1976, 63-71, (NASA TMX-73, 170).

Ettema, J. H. Blood pressure changes during mental load experiments in man. *Psychotherapy and Psychosomatics*, 1969, *17*, 191-195.

Farber, E. and Gallagher, V. Attentional demands as a measure of the influence of visibility conditions on driving task difficulty. *Highway Research Record*, 1972, *414*, 1-5.

Henry, P. H., Davis, T. Q., Engelken, E. J., Triebwasser, J. H. and Lancaster, M. C. Alcohol-induced performance decrements assessed by two link trainer tasks using experienced pilots. Aerospace Medicine, 1974, 45, 1180-1189.

Hess, R. A. and Teichgraber, W. M. Error quantization effects in compensatory tracking tasks. IEEE Transactions on Systems, Man, and Cybernetics, 1974, SMC-4, 343-349.

Hickok, J. H. Grip pressure as a measure of task difficulty in compensatory tracking tasks. Master's thesis, Naval Postgraduate School, Monterey, California, September, 1973.

Hicks, T. G. and Wierwille, W. W. Comparison of five mental workload assessment procedures in a moving-base driving simulator. Human Factors, in press.

Hilgendorf, E. L. Information processing, practice, and spare capacity. Australian Journal of Psychology, 1967, 19, 241-251.

Hoffman, E. R. and Joubert, P. N. The effect of changes in some vehicle handling variables on driver steering performance. Human Factors, 1966, 8, 245-263.

Holden, F. M., Rogers, D. B. and Replogle, C. R. Simulation of high workload operations in air to air combat. Proceedings of the AGARD Conference on Simulation and Study of High Workload Operations, AGARD-CP-146, April, 1974, A6-1 - A6-4.

Holland, M. K. and Tarlow, G. Blinking and mental load. Psychological Reports, 1972, 31, 119-127.

Hopkin, V. D. Mental workload measurement in air traffic control. Position paper prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977.

Hosman, R.J.A.W. Pilot's tracking behavior under additional workload. Delft, The Netherlands: Delft University of Technology, Department of Aeronautical Engineering, Report UTH-199, June, 1975.

Howitt, J. S. Flight-deck workload studies in civil transport aircraft. In AGARD measure of aircrew performance. Report No. N70-19780, December, 1969.

Huddleston, H. F. and Wilson, R. V. An evaluation of the usefulness of four secondary tasks in assessing the effect of a lag in simulated aircraft dynamics. Ergonomics, 1971, 14, 371-380.

Hughes, H. M., Hartman, B. O., Garcia, R. and Lozano, P. Systems simulation: A global approach to aircrew workload. Proceedings of the AGARD Conference on Simulation and Study of High Workload Operations, AGARD-CP-146, April, 1974, A1-1 - A1-14.

Jahns, D. W. Operator workload: What is it and how should it be measured? In K. D. Cross and J. J. McGrath (Eds.) Crew System Design. Santa Barbara, California: Anacapa Sciences, July, 1973. (a)

Jahns, D. W. A concept of operator workload in manual vehicle operations. Meckenheim, Germany: Forschungsinstitufuer Anthropotechnik, Report No. 14, 1973. (b)

Jenney, L. L., Older, H. J. and Cameron, B. J. Measurement of operator workload in an information processing task. Washington, D.C.: National Aeronautics and Space Administration, Contractor's Report NASA CR-2150, December, 1972.

Jennings, A. E. and Chiles, W. D. An investigation of time-sharing ability as a factor in complex performance. Human Factors, 1977, 19, 535-547.

Jex, H. R. Two applications of a critical-instability task to secondary workload research. IEEE Transactions on Human Factors in Electronics, 1967, HFE-8, 279-282.

Jex, H. R. and Allen, R. W. Research on a new human dynamic response test battery. Part I. Test development and validation. Proceedings of the 6th Annual NASA-University Conference on Manual Control, Wright-Patterson AFB, Ohio, April, 1970. (a)

Jex, H. R. and Allen, R. W. Research on a new human dynamic response test battery. Part II. Psychophysiological correlates. Proceedings of the 6th Annual NASA-University Conference on Manual Control, Wright-Patterson AFB, Ohio, April, 1970. (b)

Jex, H. R. and Clement, W. F. Defining and measuring perceptual-motor load in manual control tasks. Hawthorne, Calif.: Systems Technology, Inc., Report No. 1104-1, March, 1978.

Jex, H. R., Jewell, W. F. and Allen, R. W. Development of the dual-axis and cross-coupled critical tasks. Proceedings of the Eighth Annual Conference on Manual Control, University of Michigan, Ann Arbor, Michigan, May 1972, pp. 529-552.

Jex, H. R., McDonnell, J. D. and Phatak, A. V. A "critical" tracking task for man-machine research related to operator's effective delay time. Proceedings of the 2nd Annual NASA-University Conference on Manual Control, Massachusetts Institute of Technology, March, 1966, 361-377. (NASA-SP-128).

Johannsen, G. Nebenaufgaben als Beanspruchungsmessverfahren in Fahrzeugfuhrungsaufgaben. Zeitschrift fur Arbeitswissenschaft, 1976, 30, 45-50.

Johannsen, G. Position paper on mental workload. Prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977.

Johannsen, G., Pfendler, C. and Stein, W. Human performance and workload in simulated landing-approaches with autopilot-failures. In T. B. Sheridan and G. Johannsen (Eds.) Monitoring behavior and supervisory control. New York: Plenum, 1976, 83-95.

Jones, E. C., Jr. and Schuster, D. R. Design and development of an adaptive, auditory, and distractive stressor. IEEE Transactions on Man-Machine Systems, 1970, MMS-11/3, 161-163.

Kahneman, D., Beatty, J. and Pollack, I. Perceptual deficit during a mental task. Science, 1967, 157, 218-219.

Kahneman, D., Tursky, B., Shapiro, D. and Crider, A. Pupillary, heart rate, and skin resistance changes during a mental task. Journal of Experimental Psychology, 1969, 79, 164-167.

Kalsbeek, J.W.H. Objective measurement of mental workload: Possible applications to the flight task. Proceedings of the 55th AGARD Conference, Amsterdam, The Netherlands, 1968, 4.1 - 4.6.

Kalsbeek, J.W.H. Measurement of mental workload and of acceptable load: Possible applications in industry. International Journal of Production Research, 1969, 2, 33-45.

Kalsbeek, J.W.H. Standards of acceptable load in ATC tasks. Ergonomics, 1971, 14, 641-650.

Kalsbeek, J.W.H. Sinus arrhythmia and the dual task method in measuring mental load. In W. T. Singleton, J. G. Fox, and D. Whitfield (Eds.) Measurement of Man at Work. London: Taylor and Francis, 1973, 101-113. (a)

Kalsbeek, J.W.H. and Sykes, R. N. Objective measurement of mental load. Psychologica, 1967, 27, 253-261.

Kantowitz, B. H. and Knight, J. L., Jr. Testing tapping time-sharing, II. Auditory secondary task. Acta Psychologica, 1976, 40, 343-362.

Kantowitz, B. H. and Knight, J. L., Jr. Testing tapping time-sharing: attention demands of movement amplitude and target width. In G. E. Stelmach (Ed.) Information Processing in Motor Learning and Control. New York: Academic, 1977.

Kelley, C. R. Design applications of adaptive (self-adjusting) simulators. Proceedings of the 2nd Annual NASA-University Conference on Manual Control, Massachusetts Institute of Technology, March, 1966, 379-401. (NASA SP-128).

Kelley, C. R. and Wargo, M. J. Cross-adaptive operator loading tasks. Human Factors, 1967, 9, 395-404.

Kennedy, J. P. Time-sharing effects on pilot tracking performance. Master's thesis, Naval Postgraduate School, Monterey, California, September, 1975. (AD A016 378).

Kennedy, R. S. Two procedures for applied and experimental studies of stress. Ft. Rucker, Alabama: U.S. Army Aeromedical Research Laboratory, 70-11 NAMI 1099, February, 1970.

Kerr, B. Processing demands during mental operations. Memory and Cognition, 1973, 1, 401-412.

Kirchner, J. H. and Laurig, W. The human operator in air traffic control systems. Ergonomics, 1971, 14, 549-556.

Klein, T. J. A workload simulation model for predicting human performance requirements in the pilot-aircraft environment. Paper presented at the 14th Annual Meeting of the Human Factors Society, San Francisco, California, October 13-16, 1970.

Klein, T. J. and Cassidy, W. B. Relating operator capabilities to system demands. Dallas, Texas: LTV Aerospace, Vought Systems Division, 1972.

Klein, T. J. and Hall, A. A. An analysis of pilot performance requirements in the A-7E team. Dallas, Texas: LTV Aerospace, Vought Systems Division, VSD Report No. 2-542201 5R-5777, April, 1975.

Knowles, W. B. Operator loading tasks. Human Factors, 1963, 5, 155-161.

Kornstadt, H. J. and Pfennigsdorf, J. Evaluation of an integrated flight display for the manual IFR-landing of VTOL-aircraft. Proceedings of the AGARD Conference on Guidance and Control Displays, AGARD-CP-96, 10-1 - 10-8, 1972.

Koym, K. G. Familarity effects on task difficulty ratings. Brooks AFB, Texas: USAF Human Resources Laboratory, AFHRL-TR-77-25, June, 1977.

Kradz, M. P. The psychological stress evaluator. Ellicott City, Maryland: Howard County Police Department, Unpublished manuscript, 1974.

Kraft, C. L. and Elworth, C. L. Flight deck workload and night visual approach performance. In AGARD measure of aircrew performance. Report No. N70-19786, December, 1969.

Krause, E. F. and Roscoe, S. N. Reorganization of airplane manual flight control dynamics. W. B. Knowles, M. S. Sanders, and F. A. Muckler (Eds.) Proceedings of the Sixteenth Annual Meeting of the Human Factors Society. Santa Monica, California: Human Factors Society, 1972, 117-126.

Krebs, M. J. and Wingert, J. W. Use of the oculometer in pilot workload measurement. Washington, D.C.: National Aeronautics and Space Administration, NASA CR-144951, February, 1976.

Krebs, M. J., Wingert, J. W. and Cunningham, T. Exploration of an oculometer-based model of pilot workload. Washington, D.C.: National Aeronautics and Space Administration, NASA CR-145153, March, 1977.

Kreifeldt, J., Parkin, and Rothschild, P. Implications of a mixture of aircraft with and without traffic situation displays for air traffic management. *Proceedings of the 12th annual NASA-University Conference on Manual Control*, University of Illinois, May, 1976, 179-200. (NASA TM X-73 70).

Krivohlavy, J. Physiological correlates of the informational performance. *Activitas Nervosa Superior*, 1968, *10*, 165-171. (In Czechoslovakian). (a)

Krivohlavy, J. Pulse rate and information load during typing. *Activitas Nervosa Superior*, 1968, *10*, 172-176. (In Czechoslovakian). (b)

Krol, J. P. Variations in ATC-workload as a function of variations in cockpit workload. *Ergonomics*, 1971, *14*, 585-590.

Krzanowski, W. J. and Nicholson, A. N. Analysis of pilot assessment of workload. *Aerospace Medicine*, 1972, *43*, 993-007.

Kuhar, W. R., Gavel, P. and Moreland, J. A. Impact of automation upon air traffic control system productivity/capacity (ARTS-111). Washington, D.C.: Federal Aviation Administration, FAA-RD-77-39, November, 1976.

Lane, N. E. and Streib, M. I. The human operator simulator: Workload estimation using a simulated secondary task. Paper presented at NATO/AGARD Aerospace Medical Panel, Cologne, Germany, April, 1977.

Lane, N. E., Wherry, R. J., Jr. and Streib, M. The human operator simulator: Estimation of workload reserve using a simulated secondary task. *Proceedings of the AGARD Conference on Methods to Assess Workload*, AGARD-CPP-216, April, 1977, A11-1.

Laurell, H. and Lispar, H. A. A validation of subsidiary reaction time against detection of roadside obstacles during prolonged driving. *Ergonomics*, 1978, in press.

Laurig, W. and Phillip, U. Changes in the pulse frequency rhythm in relation to the workload. (Veranderunden der Pulsfrequenzarrhythmie in Abhangigkeit von der Arteitsschwere). *Arbeitsmdeizin Sozialmedizin, Arbeitshygiene*, 1970, *5*, 184-188. (Royal Aircraft Establishment Library Translation 1586).

Lauschner, E. A. Measurement of aircrew performance: The flight deck workload and its relation to pilot performance. AGARD Aerospace Medical Panel, AGARD-CP-56, May, 1969.

Laville, A., Teiger, C., and Duraffourg, J. An attempt to evaluate workload in a repetitive task. Paper presented at the annual conference of the Ergonomics Research Society, April, 1972.

Lebacqz, J. V. and Aiken, E. W. A flight investigation of control, display, and guidance requirements for decelerating descending VTOL instrument transitions using the X-22A variable stability aircraft. Volume I. Buffalo, New York: Calspan Corporation, Ak-5336-F-1, September, 1975.

Leplat, J. Factors determining workload. *Ergonomics*, 1978, *21*, 143-149.

Leplat, J. and Pailhous, J. The analysis and evaluation of mental work. In W. T. Singleton, J. G. Fox, and D. Whitfield (Eds.) *Measurement of man at work*. London: Taylor and Francis, 1973, 51-56.

Levine, J. M., Ogden, G. D. and Eisner, E. J. Measurement of workload by secondary tasks. Washington, D.C.: Advanced Research Resources Organization, Contract No. NAS2-9637, January, 1978.

Levison, W. H. A model for task interference. *Proceedings of the 6th Annual NASA-University Conference on Manual Control*, Wright-Patterson AFB, Ohio, April 7-9, 1970, 585-616.

Levison, W. H. Position paper prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977.

Lindquist, O. H. Design implications of a better view of the multichannel capacity of a pilot. *Proceedings of the AGARD Conference on Guidance and Control Displays*, AGARD-CP-96, 5-1 - 5-6, February, 1972.

Linn, V. C., Jr. The parotid fluid technique for the evaluation of mental stress in a production situation. Texarkana, Texas: US Army Logistics Management Center, USAMC Intern Training Center, USAMC-ITC Report No. 2-72-05, July, 1972.

Linton, P. M. VFA-V/STOL crew loading analysis. Warminster, Pennsylvania: U.S. Naval Air Development Center, NADC-75209-40, May, 1975.

Linton, P. M., Jahns, D. W. and Chatelier, P. R. Operator workload assessment model: An evaluation of a VF/VA-V/STOL system. *Proceedings of the AGARD Conference on Methods to Assess Workload*, AGARD-CPP-216, April, 1977, A12-1 - A12-11.

Lisper, H. O., Laurell, H. and Stening, G. Effects of experience of the driver on heart-rate, respiration-rate, and subsidiary reaction time in a three hours continuous driving task. *Ergonomics*, 1973, *16*, 501-506.

Lovesey, E. J. Inflight recording of helicopter pilot activity. Proceedings of AGARD Conference on Studies on Pilot Workload, AGARD-CPP-217, April, 1977, B3-1 - B3-10.

Lucsak, H. and Laurig, W. An analysis of heart rate variability. Ergonomics, 1973, 16, 85-97.

Machac, M. Mental load, fatigue, and recovering. Psychologic, 1971, 6, 72-79. (In Czechoslovakian).

Mashhour, M. The effect of motion on attention in man-machine systems. Stockholm, Sweden: University of Stockholm, The Psychological Laboratories, April, 1969.

McDonald, L. B. and Ellis, N. C. Stress threshold for drivers under various combinations of discrete and tracking workload. Proceedings of the 19th annual meeting of the Human Factors Society, Dallas, Texas, October, 1975, 488-493. (a)

McDonald, L. B. and Ellis, N. C. Driver workload for various turn radii and speeds. In Driver performance studies: Transportation Research Record 530. Washington, D.C.: Transportation Research Board, TRR 530, 1975, 18-30. (b)

McFeely, T. E. Pupil diameter and the cross-adaptive critical tracking task; A method of workload measurement. Master's thesis, Naval Post-graduate School, Monterey, California, June, 1972. (AD 749 075).

McGrath, J. J. Temporal orientation and task performance. Goleta, California: Human Factors Research, 719-IC, January, 1969. (AD 758 909).

McHugh, W. B., Brictson, C. A. and Naitoh, P. Emotional and biochemical effects of high workload. Proceedings of the AGARD Conference on Simulation and Study of High Workload Operations, AGARD-CP-146, April, 1974, A12-1 - A12-9.

McLean, J. R. and Hoffmann, E. R. Steering reversals as a measure of driver performance and steering task difficulty. Human Factors, 1975, 17, 248-256.

Merhav, S. J. and Ya'acov, O. B. Control augmentation and workload reduction by kinesthetic information from the manipulator. Proceedings of the 12th Annual NASA-University Conference on Manual Control, University of Illinois, May, 1976. (NASA TM X-73 70).

Michon, J. A. A note on the measurement of perceptual motor load. Ergonomics, 1964, 7, 461-463.

Michon, J. A. Tapping regularity as a measure of perceptual motor load. Ergonomics, 1966, 9, 401-412.

Michon, J. A. and Doorne, H. van Equipment note: A semi-portable apparatus for the measurement of perceptual motor load. Ergonomics, 1967, 10, 67-22.

Miller, R. G. and Rubin, R. T. Stress of aircraft carrier landings. 1. Corticosteroid responses in Naval aviators. San Diego, California: Navy Medical Neuropsychiatric Research Unit, Report No. NMNRU-70-16, April, 1970.

Mobbs, R. F., David, G. C. and Thomas, J. M. An evaluation of the use of heart rate irregularity as a measure of mental workload in the steel industry. London, England: British Steel Corporation, EISRA, OR/HR/25/71, August, 1971.

Monty, R. A. and Ruby, W. J. Effects of added workload on compensatory tracking for maxium terrain following. Human Factors, 1965, 7, 207-214.

Moray, N. Mental workload position paper. Position paper prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977.

Morgan, T. R. Inflight physiological data acquistion system. Brooks AFB, Texas: USAF School of Aerospace Medicine, SAM-TR-75-46, December, 1975.

Morrissette, J. O., Crannell, C. W. and Switzer, S. A. Group performance under various conditions of workload and information redundancy. Wright-Patterson AFB, Ohio: Aerospace Medical Research Laboratory, AMRL-TR-65-16, April, 1965.

Mulder, G. The heart of mental effort. Position paper prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977.

Mulder, G. and Mulder-Hajonides van der Meulen, W.R.E.H. Mental load and the measurement of heart rate variability. Ergonomics, 1973, 16, 69-83.

Murphy, J. V. and Gurman, B. S. The integrated cockpit procedure for identifying control and display requirements of aircraft in advanced time periods. Proceedings of the AGARD Conference on Guidance and Control Displays, AGARD-CP-96, 4-1 - 4-7, 1972.

Murphy. M. R. Coordinated crew performance in commercial aircraft operations. Proceedings of the 21st Annual Meeting of the Human Factors Society, San Francisco, California, October, 1978, 416-420.

Murphy, M. R., McGee, L. A., Palmer, E. A., Paulk, C. H. and Wempe, T. E. Simulator evaluation of three situation and guidance displays for V/STOL zero/zero landings. Proceedings of the 10th Annual NASA-University Conference on Manual Control, Wright-Patterson AFB, Ohio, April 9-11, 1974.

Murrell, J. F. Pilot's assessment of their cockpit environment. In *Problems of the Cockpit Environment*, AGARD Conference Proceedings No. 55, March, 1970.

Nagaraja Rao, B. K. and Griffin, J. J. Secondary task performance of helicopter pilots during low-level flight. University of Southampton, Institute of Sound and Vibration Research, Report No. ISVR-TR-54, December, 1971.

Navon, D. and Gopher, D. On the economy of the human processing system: A model of multiple capacity. Haifa, Israel: Technion Technical Report AFOSR-77-1, 1977.

Nicholson, A. N. Aircrew workload during the approach and landing. *Aeronautical Journal*, 1973, *77*, 286-289

Nicholson, A. N., Hill, L. E., Borland, R. G., and Krzanowski, W. J. Influence of workload on the neurological state of a pilot during the approach and landing. *Aerospace Medicine*, 1973, *44*, 146-152.

Noble, M. and Trumbo, D. The organization of skilled response. *Organizational Behavior and Human Performance*, 1967, *2*, 1-25.

Noel, C. E. Pupil diameter versus task layout. Master's thesis, Naval Postgraduate School, Monterey, California, September, 1974.

North, R. A. Task components and demands as factors in dual-task performance. Savoy, Illinois: University of Illinois at Urgana-Champaign, ARL-77-2/AFOSR-77-2, January, 1977.

North, R. A. and Goper, D. Measures of attention as predictors of flight performance. *Human Factors*, 1976, *18*, 1-14.

Noyer, A. Mental fatigue and palmar skin resistance. *Travail Humain*, 1971, *34*, 289-298. (In French).

O'Connor, M. F. and Buede, B. M. The application of a decision analytic techniques to the test and evaluation phase of the acquisition of a major air system. McLean, Virginia: Decisions and Designs, Technical Report 77-3, April, 1977.

O'Donnell, R. D. Secondary task assessment of cognitive workload in alternative cockpit configurations. Wright-Patterson AFB, Ohio: Aerospace Medical Research Laboratory, AMRL-TR-75-49, 1975. (AGARD-CPP-181).

O'Donnell, R. D. Secondary task assessment of cognitive workload in alternative cockpit configurations. In *Higher mental functioning in operational environments*. North Atlantic Treaty Organization, April, 1976.

O'Donnell, R. D. and Spicuzza, R. J. Pilot performance assessment in systems using integrated digital avionics. *Proceedings of the 46th Annual Meeting of the Aerospace Medical Association*, San Francisco, California, 1975.

Ohhara, S. Changes of tracking performance, respiration, and heart rate during experimentally induced anxiety. *Japan Air Self Defense Force, Aeromedical Laboratory Reports*, 1970, *11*, 198-205. (In Japanese).

Older, H. J. and Jenney, L. L. Psychological stress measurement through voice output analysis. Alexandria, Virginia: The Planar Corporation, Contract NASA 9-14146, March, 197[illegible]

Olson, B. A. Display and control requirements study for a V/STOL tactical aircraft. Wright-Patterson AFB, Ohio: USAF Flight Dynamics Laboratory, AFFDL-TR-66-114, December, 1966.

Onstott, E. D. Task interference in multi-axis aircraft stabilization. *Proceedings of the 12th Annual NASA-University Conference on Manual Control*, University of Illinois, May, 1976, 80-103. (NASA TM X-73, 70).

Onstott, E. D. and Faulkner, W. H. Predication of pilot reserve attention capacity during air-to-air target tracking. *Proceedings of the 13th Annual NASA-University Conference on Manual Control*, Massachusetts Institute of Technology, June 15-17, 1977, 136-142.

Opmeer, C.H.J.M. The information content of successive RR-interval times in the ECG. Preliminary results using factor analysis and frequency analysis. *Ergonomics*, 1973, *16*, 105-112.

Parks, D. L. Current workload methods and emerging challenges. Seattle, Wash.: The Boeing Co., Document No. D6-44563TN, July, 1977.

Parks, D. L. and Springer, W. E. Human factors engineering analytic process definition and criterion development for Computer Aided Function-allocation Evaluation System (CAFES). Seattle, Washington: Boeing Aerospace Company, D180-18750-1, January, 1976.

Pettyjohn, F. S., McNeil, R. J., Akers, L. A. and Faber, J. M. Use of inspiratory minute volumes in evaluation of rotary and fixed wing pilot workload. *Proceedings of the AGARD Conference on Methods to Assess Workload*, AGARD-CPP-216, April, 1977, A9-1 - A9-2. (a)

Pettyjohn, F. S., McNeil, R. J., Akers, L. A. and Faber, J. M. Use of inspiratory minute volumes in evaluation of rotary and fixed wing pilot workload. Fort Rucker, Alabama: U.S. Army Aeromedical Research Laboratory, USAARL Report No. 77-9, April, 1977. (b)

Pew, R. W. Position paper on workload. Prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977.

Phatak, A. V. Improvement in weapon system effectiveness by application of identification methods for determining human operator performance decrements under stress conditions. Palo Alto, California: Systems Control, December, 1973.

Philipp, U., Reiche, D. and Kirchner, J. H. The use of subjective rating. *Ergonomics*, 1971, *14*, 611-616.

Phillips, J. F. The feasibility of short interval time estimation as a methodology to forecast human performance of a specified task. Red River Army Depot, Texarkana, Texas: DARCOM Intern Training Center, DARCOM-ITC-02-08-76-010, April, 1976.

Poston, A. M. A survey of existing computer programs for aircrew workload assessment. Aberdeen Proving Ground, Maryland: U.S. Army Human Engineering Laboratory, Technical Memorandum 13-78, May, 1978.

Potempa, K. W. A catalog of human-factors techniques for testing new systems. Wright-Patterson AFB, Ohio: USAF Human Resources Laboratory, AFHRL-TR-68-15, February, 1969.

Price, D. L. The effects of certain gimbal orders on target acquisition and workload. *Human Factors*, 1975, *17*, 571-576.

Price, H. E. Development of potential roles of supersonic transport crews. Chatsworth, California: Serendipity Associates, TR 20-66-3, December, 1965.

Price, H. E., Honsberger, W. D., and Ereneta, W. J. A study of potential roles of supersonic transport crews and some implications for the flight deck, Volume I: Workload, crew roles, flight deck concepts, and conclusions. Moffett Field, California: National Aeronautics NASA CR-561, October, 1966.

Pritsker, A.A.B., Wortman, D. R., Seum, C. S., Chubb, G. P. and Siefert, D. J. SAINT: Volume I. Systems analysis of integrated network of tasks. Wright-Patterson AFB, Ohio: Aerospace Medical Research Laboratory, AMRL-TR-73-126, April, 1974.

Rasmussen, J. Reflections on the concept of perator workload. Position paper prepared for NATO Symposium on Mental Workload. Mati, Greece September, 1977.

Reising, J. M. The definition and measurement of pilot workload. Wright-Patterson AFB, Ohio: USAF Flight Dynamics Laboratory, AFFDL-TM-72-4-FGR, February, 1977.

Repa, B. S. and Wierwille, W. W. Driver performance in controlling a driving simulator with varying vehicle response characteristics. SAE Paper No. 760779, October, 1976.

pko, J. D., Loeb, M. and Brown, B. R. Behavioral effects of prolonged exposure to continuous and intermittent noise. Louisville, Kentucky: University of Louisville, Performance Research Laboratory, TR-74-29, June, 1974.

Replogle, C. R., Holden, F. M., Gold, R. E., Kalak, L. L., Jonas, F. and Potor, C., Jr. Human operator performance in hypoxic stress. Wright-Patterson AFB, Ohio: USAF Aerospace Medical Research Laboratory, AMRL-TR-71-29, Paper No. 31, December, 1971.

Rohmert, W. An international symposium on objective assessment of workload in air traffic control tasks: Held at the Institute of Arbeitswissenschaft, The University of Technology, Darmstadt, German Federal Republic. *Ergonomics*, 1971, *14*, 545-547.

Rohmert, W. Determination of stress and strain of air traffic control officers. *Proceedings of the AGARD Cor nce on Methods to Assess Workload*, AGARD-CPP-216, April, 1977, A6-1 - A6-8.

Rohmert, W., Laurig, W., Philipp, U. and Luczak, H. Heart rate variability and workload measurement. *Ergonomics*, 1973, *16*, 33-44.

Rolfe, J. M. Multiple task performance: Operator overload. *Occupational Psychology*, 1971, *45*, 125-132.

Rolfe, J. M. Whither workload. *Applied Ergonomics*, 1973, *4*, 8-10. (a)

Rolfe, J. M secondary task as a measure of mental load. In W. T. Singleton, J. G. Fox, and D. Whitfield .) *Measurement of Man at Work*. London: Taylor and Francis, 1973, 135-148. (b)

Rolfe, J. M., Chappelow, J. W., Evans, R. L., Lindsay, S.J.E. and Browning, A. C. Evaluating measures of workload using a flight simulator. *Proceedings of the AGARD Conference on Simulation and Study of High Workload Operations*, AGARD-CP-146, April, 1974, A4-1 - A4-13.

Rolfe, J. M. and Lindsay, S.J.E. Flight deck environment and pilot workload: Biological measures of workload. *Applied Ergonomics*, 1973, *4*, 199-206.

Rosch, E. and Wempe, T. Secondary task for full flight simulation incorporating tasks that commonly cause pilot error: Time estimation Moffett Field, California: NASA-AMES Research Center, NASA-TM-X-74153, October, 1975.

Roscoe, A. H. Pilot workload during steep gradient approaches. Farnsborough, England: Royal Aircraft Establishment, Flight Systems Department, Technical Memorandum No. TM FS 78, 1976.

Roscoe, A. H. (Ed.) *Assessing pilot workload*. AGARD-AG-233, February, 1978. (a)

Roscoe, A. H. Physiological methods. In A. H. Roscoe (Ed.) *Assessing pilot workload*. AGARD-AG-233, February, 1978 (b).

Roscoe, S. N. Assessment of pilotage error in airborne area navigation procedures. *Human Factors*, 1974, *16*, 223-228.

Rosenbrock, F. Hardware problems in ergonomics measurements. *Ergonomics*, 1971, *14*, 617-623.

Roult, A. Outlines of a position paper. Position paper prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977.

Rouse, W. B. Approaches to mental workload. Position paper prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977. (a)

Rouse, W. B. Human-computer interaction in multitask situations. *IEEE Transactions on Systems, Man, and Cybernetics*, 1977, *SMC-7*, 384-392. (b)

Sanders, A. F. Some remarks on mental load. Position paper prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977.

Sanders, M. S., Jankovich, J. J. and Goodpaster, P. R. Task analysis for the jobs of train conductor and brakeman. Crane, Indiana: Naval Ammunition Depot, RDTR-No. 623, July, 1974.

Sanders, M. G., Simmons, R. R., Hofmann, M. A. and DeBonis, J. N. Visual workload of the co-pilot/ navigator during terrain flight. *Proceedings of the Human Factors Twenty-First Annual Meeting*. San Francisco, California: Human Factors Society, October 1977, 262-266.

Savage, R. E., Wierwille, W. W. and Cordes, R. E. Evaluating the sensitivity of various measures of operator workload using random digits as a secondary task. *Human Factors*, in press.

Sayers, B. M. Analysis of heart rate variability. *Ergonomics*, 1973, *16*, 17-32.

Schiffler, R. J., Geiselhart, R. and Ivey, L. Crew composition study for an Advanced Tanker/Cargo Aircraft (ATCA). Wright-Patterson, AFB, Ohio: USAF Aeronautical Systems Division, ASD-TR-76-20, October, 1976.

Schiflett, S. G. Operator workload: An annotated bibliography. Patuxent River, Maryland: US Naval Air Test Center, SY-257R-76, December, 1976.

Schori, T. R. A comparison of visual, auditory, and cutaneous tracking displays when divided attention is required to a cross-adaptive loading task. *Ergonomics*, 1973, *16*, 153-158.

Schori, T. R. and Jones, B. W. Smoking and workload. *Journal of Motor Behavior*, 1975, *7*, 113-120.

Schouten, J. F., Kalsbeek, J.W.H., and Leopold, F. F. On the evaluation of perceptual and mental load. *Ergonomics*, 1962, *5*, 251-260.

Schultz, W. C., Newell, F. D. and Whitbeck, R. F. A study of relationships between aircraft system performance and pilot ratings. *Proceedings of the 6th Annual NASA-University Conference on Manual Control*, Wright-Patterson AFB, Ohio, April 7-9, 1970, 339-340.

Schwartz, J. J. and Ekkers, C. L. Estimation of task loading by observing and regulating complex technical systems. *Mens en onderneming*, 1976, *76*, 85-108. (In Dutch).

Seibel, R., Christ, R. E. and Teichner, W. H. Perception and short term memory under workload stress. Port Washington, New York: U.S. Naval Training Device Center, NAVTRADEVCEN 1303-2, June, 1964.

Seifert, R., Daniels, A. F. and Schmidt, K. A method of man-display/control system evaluation. *Proceedings of the AGARD Conference on Guidance and Control Displays*, AGARD-CP-96, 8-1 - 8-8.

Senders, J. W. The human operator as a monitor and controller of multidegree of freedom systems. *IEEE Transactions on Human Factors in Electronics*, 1964, *HFE-5*, 2-5.

Senders, J. W. The estimation of operator workload in complex systems. In K. B. DeGreene (Ed.) *Systems Psychology*. New York: McGraw-hill, 1970.

Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W. and Ward, J. L. The attentional demand of automobile driving. *Highway Research Record*, 1967, No. 195, 15-33.

Sheridan, T. B. and Stassen, H. G. Definitions, models and measures of human workload. Position paper prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977.

Sherman, M. R. The relationship of eye behavior, cardiac activity and electromyographic responses to subjective reports of mental fatigue and performance on a Doppler identification task. Master's thesis, Naval Postgraduate School, Monterey, California, September, 1973. (AD 769 754).

Shulman, H. G. and Briggs, G. E. Studies of performance in complex aircrew tasks. Columbus, Ohio: The Ohio State University, Research Foundation, RF Project 2718, Final Report, December, 1971.

Siegel, A. I., Lanterman, R. S., Platzer, H. L. and Wolf, J. J. Techniques for evaluating operator loading in man-machine systems: Development of a method for real time assessment of operator overloading. Wayne, Pennsylvania: Applied Psychological Services, January, 1976.

Siegel, A. I. and Williams, A. R., Jr. Identification and measurement of intellective load carrying thresholds. Wayne, Pennsylvania: Applied Psychological Services, AFOSR-TR-75-0593, December, 1974.

Siegel, A. I. and Wolf, J. J. Man-machine simulation models: Psychosocial and performance interaction. New York: Wiley, 1969.

Siegel, A. I., Wolf, J. J., Fischl, M. A., Miehle, W. and Chubb, G. P. Modification of the Siegel-Wolf operator simulation model for on-line experimentation. Wright-Patterson AFB, Ohio: USAF Aerospace Medical Research Laboratory, AMRL-TR-71-60, June, 1971.

Siegel, A. I., Wolf, J. J. and Sorenson, R. T. Technicians for evaluating operator loading in man-machine system: Evaluation of a one or a two-operator system evaluating model through a controlled laboratory test. Wayne, Pennsylvania: Applied Psychological Services, Contract Nonr 2-492(00), July, 1962. (AD 284 182)

Simmons, R. R. and Kimball, K. A. Methodological considerations of visual workloads of helicopter pilots. Proceedings of the AGARD Conference on Methods to Assess Workload, AGARD-CPP-216, April, 1977, A2-1.

Simmons, R. R., Kimball, K. A. and Diaz, J. J. Measurement of aviator visual performance and workload during helicopter operations. Ft. Rucker, Alabama: U.S. Army Aeromedical Research Laboratory, 77-4. December, 1976.

Simonov, P. V. and Frolov, M. V. Analysis of the human voice as a method of controlling emotional state: Achievements and goals. Aviation, Space, and Environmental Medicine, 1977, 48, 23-25.

Simpson, C. A. and Hart, S. G. Required attention for synthesized speech perception for two levels of linguistic redundancy. Paper presented at the 93rd meeting of the Acoustical Society of America, State College, Pennsylvania, June 7-10, 1977.

Simpson, C. G. Improved displays and stabilization in general aviation aircraft. Moffett Field, California National Aeronautical and Space Administration, N69-24238, November, 1968. (AGARD Symposium paper).

Sinaiko, H. W. Third international congress on ergonomics. London, England: Office of Naval Research, ONRL-C-19-67, November, 1967.

Smith, W. S., Jr. Effects of neuromuscular tension in the use of an isometric hand controller. Monterey, California: U.S. Naval Post-graduate School, Master's thesis, December, 1972.

Soede, M. Reduced mental capacity and behavior of a rider of a bicycle simulator under alchohol stress or under dual task load. Proceedings of the 13th Annual NASA-University Conference on Manual Control, Massachusetts Institute of Technology, June 15-17, 1977, 143-151. (a)

Soede, M. On mental load and reduced mental capacity; some considerations concerning laboratory and field investigations. Position paper prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977. (b)

Soliday, S. M. Effects of task loading on pilot performance during simulated low-altitude high-speed flight. Fort Eustis, Virginia: U.S. Army Transportation Research Center, USATRECOM 64-69, February, 1965.

Soliday, S. M. and Schohan, B. Task loading of pilots in simulated low-altitude high-speed flight. Human Factors, 1965, 7, 45-53.

Soutendam, J. Instruments and methodology for the assessment of physiological cost of performance in stressful continuous operations the air traffic services tower environment. Proceedings of the AGARD Conference on Methods to Assess Workload, AGARD-CPP-216, April, 1977, A7-1 - A7-32.

Sperandio, J. C. Variation of operator's strategies and regulating effects on workload. Ergonomics, 1971, 14, 571-577.

Sperandio, J. The regulation of working methods as a function of workload among air traffic controllers. Ergonomics, 1978, 21, 195-202.

Spicuzza, R. J., Pinkus, A. R. and O'Donnell, R. D. Development of performance assessment methodology for the digital avionics information system. Dayton, Ohio: Systems Research Laboratories Final Report, 1 December 1973 - 30 June 1974, August, 1974.

Spyker, D. A., Stackhouse, S. P., Khalafalla, A. S. and McLane, R. C. Development of techniques for measuring pilot workload. Washington D.C.: National Aeronautics and Space Administration, Contractor's Report NASA CR-1888, November, 1971.

Stackhouse, S. Workload evaluation of LLNO display. Minneapolis, Minnesota: Honeywell, 7201-3408, October, 1973.

Stackhouse, S. P. The measurement of pilot workload in manual control system. Minneapolis, Minnesota: Honeywell, Inc., FP398 FR1, January, 1976.

Stackhouse, S. P. Measurement of aircrew information processing workload. Minneapolis, Minnesota: Honeywell, Unpublished manuscript, 1978.

Stager, P. and Muter, P. Instructions and information processing in a complex task. Journal of Experimental Psychology, 1971, 87, 291-294.

Stager, P. and Zufelt, K. Dual-task method in determining load differences. Journal of Experimental Psychology, 1972, 94, 113-115.

Stamford, B. A. Validity and reliability of subjective rating of perceived exertion during work. Ergonomics, 1976, 19, 53-60.

Steininger, K. Subjective ratings of flying qualities and pilot workload in the operation of a short haul jet transport aircraft. Proceedings of AGARD Conference on Studies on Pilot Workload, AGARD-CPP-217, April, 1977, B11-1.

Steininger, K. and Wistuba, C. Minimum flight crew of transport aircraft. Methods for measuring workload of flight crews. Hamburg, West Germany: Deutsche Forschungs-and Versuchsanstatt fuer Luft-und Raunfahrt, Report No. DLR-IB-355-74/3, 1974. (In German).

Stephens, B. W. and Michaels, R. M. Time-sharing between two driving tasks: Simulated steering and recognition of road signs. Paper presented at the 43rd Annual Meeting of the Highway Research Board, Washington, D.C., January, 1964.

Sternberg, S. High-speed scanning in human memory. In R. N. Haber (Ed.) New York: Holt, Rinehart and Winston, 1969.

Storm, W. F. and Hapenney, J. D. Mission-Crew fatigue during rivet joint operations. Brooks AFB, Texas: USAF School of Aerospace Medicine, SAM-TR-76-36, September, 1976.

Storm, W. F., Hartman, B. O., Intano, G. P. and Peters, G. L. Endocrine-metabolic effects in short-duartion high-workload missions: Feasibility study. Brooks AFB, Texas: USAF School of Aerospace Medicine. SAM-TR-76-30, August, 1976.

Street, R. L., Singh, H. and Hale, P. N., Jr. The evaluation of mental stress through the analysis of parotid fluid. Human Factors, 1970, 12, 453-455.

Strieb, M. I. The human operator simulator volume I: Introduction and overview. Willow Grove, Pennsylvania: Analytics, August, 1975.

Strieb, M. I., Glenn, F. A., Fisher, C. and Fitts, L. B. Chapter VII from the human operator simulator volume VII. LAMPS air tactical officer simulation. Willow Grove, Pennsylvania: Analytics, November, 1976.

Strother, D. D. Aircrew performance in army aviation. Proceedings of Conference 27-29 November 1973, U.S. Army Aviation Center, Fort Rucker, Alabama, November, 1973, 188-192.

Strother, D. D. Visual and manual workload of the helicopter pilot. Paper presented at the Annual National Forum of the American Helicopter Society, Washington, D.C., May, 1974. (Preprint No. 821).

Sun, P. B., Keane, W. P. and Stackhouse, S. P. The measurement of pilot workload in manual control systems. Proceedings of Aviation Electronics Symposium, Fort Monmouth, New Jersey, April, 1976.

Teichgraber, W. M. The effects of signal quantization on compensatory tracking performance. Unpublished master's thesis, Naval Postgraduate School, December, 1972.

Teiger, C. Regulation of activity: an analytical tool for studying workload in perceptual motor tasks. Ergonomics, 1978, 21, 203-213.

Terbraak, F. High workload tasks of aircrew in the tactical strike, attack, and reconnaissance roles. In AGARD simulation and study of high workload operations, AGARD-CP-146, October, 1974.

Thorne, R. G. Pilot workload: A conceptual model. AGARD Conference Proceedings No. 119 on Stability and Control, Braunschweig, Germany, April 10-13, 1972, 21-1-21-6. (AGARD-CP-119).

Triggs, T. J. Aspects of mental workload. Proceedings of the IEEE-GMMS ERS International Symposium on Man-Machine Systems, Cambrige, England, September, 1969.

Trumbo, D. and Noble, M. Response uncertainty in dual-task performance. Organizational Behavior and Human Performance, 1972, 7, 203-215.

Trumbo, D., Noble, M. and Swink, J. Secondary task interference in the performance of tracking tasks. Journal of Experimental Psychology, 1967, 73, 232-240.

Ursin, H. and Ursin, R. Physiological indicators of mental load. Position paper prepared for NATO Symposium on Mental Workload. Mati, Greece, September, 1977.

van Gigch, J. P. A model for measuring the information processing rates and mental load of complex activities. Canadian Operational Research Society Journal, 1970, 8, 116-128. (a)

van Gigch, J. P. Applications of a model used in calculating the mental load of workers in industry. Canadian Operational Research Society Journal, 1970, 8 176-184. (a)

Verplank, W. L. Is there an optimum workload in manual control? Proceedings of the 12th Annual NASA-University Conference on Manual Control, University of Illinois, May, 1976, 72-79. (NASA TM X-73, 70).

Waller, M. C. An investigation of correlation between pilot scanning behavior and workload using stepwise regression analysis. Hampton, Virginia: NASA Langley Research Center, NASA TM X-3344, March, 1976.

Watson, B. L. The effect of secondary tasks on pilot describing functions in a compensatory tracking task. Toronto, Canada: University of Toronto, Institute for Aerospace Studies, UTIAS Technical Note No. 178, June, 1972.

Waugh, J. D. Pilot performance in helicopter simulator. Aberdeen Proving Ground, Maryland: U.S. Army Engineering Laboratory, Technical Memorandum 23-75, September, 1975.

Weir, D. H. and Klein, R. H. Measurement and analysis of pilot scanning behavior during simulated instrument approaches. Proceedings of the 6th Annual NASA-University Conference on Manual Control, Wright-Patterson AFB, April, 1970, 83-108.

Welford, A. T. Mental workload as a function of demand, capacity, strategy and skill. Ergonomics, 1978, 21, 151-167.

Wempe, T. E. and Baty, D. L. Human information processing rates during certain, multiaxis tracking tasks with a concurrent auditory task. IEEE Transactions on Man-Machine Systems, 1968, MMS-9, 129-138.

Westbrook, C. B., Anderson, R. O. and Pietrzak, P. E. Handling qualities and pilot workload. Wright-Patterson AFB, Ohio: AF Flight Dynamics Laboratory, AFFDL-FDCC-TM-66-5, September, 1966.

Weverinke, P. H. Human control and monitoring-models and experiments. Proceedings of the 12th Annual NASA-University Conference on Manual Control, University of Illinois, May, 1976, 14-38. (NASA TMX-73, 170).

Weverinke, P. H. Performance and workload analysis of inflight helicopter tasks. Proceedings of the 13th Annual NASA-University Conference on Manual Control, Massachusetts Institute of Technology, June 15-17, 1977, 106-117.

Weverinke, P. H. and Smit, J. A simulator study to investigate human operator workload. Proceedings of the AGARD Conference on Simulation and Study of High Workload Operations, AGARD-CP-146, April, 1974, A2-1 - A2-6.

White, R. T. Task analysis methods: Review and development of techniques for analyzing mental workload in multiple-task situations. St. Louis, Missouri: McDonnell Douglas Corporation, MDC J 5291, September, 1971.

White, R. T. Mental workload assessment, I. Laboratory investigation of decision-making and short-term memory in multiple-task situation. St. Louis, Missouri: McDonnell Dougals, MDC J6662/01, December, 1975.

White, R. T. and Gaume, J. G. Mental workload assessment, III. Laboratory evaluation of one subjective and two physiological measures of mental workload. St. Louis, Missouri: McDonnell Douglas Corporation, Report MDC J7024/01, December, 1975.

White, R. T. and Ware, C. T. Prediction of human operator performance in the design of command and control systems. Paper presented at the Annual Meeting of the Western Psychological Association, Vancouver, British Columbia, June 21, 1969. (Douglas Paper 5539).

Wickens, C. D. The effect of time-sharing on the performance of information processing tasks: A feedback control analysis. Ann Arbor, Michigan: The University of Michigan, Human Performance Center, Technical Report No. 51, August, 1974.

Wickens, C. D. The effects of divided attention on information processing in manual tracking. Journal of Experimental Psychology, 1976, 2, 1-13.

Wickens, C. D. Position paper. Paper presented at NATO Conference on Workload. Matapan, Greece, September, 1977.

Wickens, C. D. and Gopher, D. Control theory measures of tracking as indices of attention allocation strategies. Human Factors, 1977, 19, 349-365.

Wickens, C. D., Isreal, J., McCarthy, G., Gopher, D. and Donchin, E. The use of event-related potentials in the enhancement of system performance. Proceedings of the 12th Annual NASA-University Conference on Manual Control, University of Illinois, May, 1976, 124-134. (NASA TM X-73, 70).

Wickens, C. D., Isreal, J. and Donchin, E. The event related cortical potential as an index of task workload. Proceedings of the 21st Annual Meeting of the Human Factors Society, San Francisco, California, 1977.

Wickens, C. D. and Kessel, C. The effects of participatory mode and task workload on the detection of dynamic system failures. Proceedings of the 13th Annual NASA-University Conference on Manual Control, Massachusetts Institute of Technology, June 15-17, 1977, 126-135.

Wierwille, W. W. and Gutmann, J. C. Comparison of primary and secondary task measures as a function of simulated vehicle dynamics and driving conditions. Human Factors, 1978, 20, 233-244.

Wierwille, W. W., Gutmann, J. C., Hicks, T. G., and Muto, W. H. Secondary task measurement of workload as a function of simulated vehicle dynamics and driving conditions. Human Factors, 1977, 19, 557-565.

Wierwille, W. W. and Williges, R. C. Survey and analysis of operator workload assessment techniques. Systemetrics, Inc. Rept. S-78-101, September, 1978. (Navair Contract N00421-77-C-0083; AD-A059501).

Wildervanck, C., Mulder, G. and Michon, J. A. Mapping mental load in car driving. Ergonomics, 1978, 21, 225-229.

Wingert, J. W. Function interlace modifications to analytic workload prediction. In K. D. Cross and J. J. McGrath (Eds.) Crew System Design. Santa Barbara, California: Anacapa Sciences, July, 1973.

Wisner, A. Electrophysiological measures for tasks of low energy expenditure. In W. T. Singleton, J. G. Fox and D. Whitfield (Eds.) Measurement of man at work. London: Taylor and Francis, 1973, 61-73.

Wolf, J. D. Workload evaluation of control-display configurations for approach to landing. Minneapolis, Minnesota: Honeywell, SRC Report F0548-IR, August, 1977.

TABLE 1

Classification of Universal Operator Behavior Dimension
(After Berliner, Angell, and Shearer, 1964)

Processes	Activities	Specific Behaviors
1. Perceptual processes	1.1 Searching for and receiving information	1.1.1 Detects
		1.1.2 Inspects
		1.1.3 Observes
		1.1.4 Reads
		1.1.5 Receives
		1.1.6 Scans
		1.1.7 Surveys
	1.2 Identifying objects, actions, events	1.2.1 Discriminates
		1.2.2 Identifies
		1.2.3 Locates
2. Mediational processes	2.1 Information processing	2.1.1 Categorizes
		2.1.2 Calculates
		2.1.3 Codes
		2.1.4 Computes
		2.1.5 Interpolates
		2.1.6 Itemizes
		2.1.7 Tabulates
		2.1.8 Translates
	2.2 Problem solving and decision-making	2.2.1 Analyzes
		2.2.2 Calculates
		2.2.3 Chooses
		2.2.4 Compares
		2.2.5 Computes
		2.2.6 Estimates
		2.2.7 Plans
3. Communication processes		3.1 Advises
		3.2 Answers
		3.3 Communicates
		3.4 Directs
		3.5 Indicates
		3.6 Informs
		3.7 Instructs
		3.8 Requests
		3.9 Transmits
4. Motor processes	4.1 Simple/Discrete	4.1.1 Activates
		4.1.2 Closes
		4.1.3 Connects
		4.1.4 Disconnects
		4.1.5 Joins
		4.1.6 Moves
		4.1.7 Presses
		4.1.8 Sets
	4.2 Complex/Continuous	4.2.1 Adjusts
		4.2.2 Aligns
		4.2.3 Regulates
		4.2.4 Synchronizes
		4.2.5 Tracks

TABLE 2

Classification of Workload Methodologies Dimension

- 1. Subjective Opinion
  - 1.1 Rating Scales
  - 1.2 Interviews and Questionnaires
- 2. Spare Mental Capacity
  - 2.1 Task Analytic
    - 2.1.1 Task Component, Time Summation
    - 2.1.2 Information-Theoretic
  - 2.2 Secondary Task
    - 2.2.1 Nonadaptive, Arithmetic/Logic
    - 2.2.2 Nonadaptive, Tracking
    - 2.2.3 Time Estimation
    - 2.2.4 Adaptive, Arithmetic/Logic
    - 2.2.5 Adaptive, Tracking
  - 2.3 Occlusion
- 3. Primary Task
  - 3.1 Single Measures
  - 3.2 Multiple Measures
  - 3.3 Math Modeling
- 4. Physiological Measures
  - 4.1 Single Measures
    - 4.1.1 FFF
    - 4.1.2 GSR
    - 4.1.3 EKG
    - 4.1.4 EMG
    - 4.1.5 EEG
    - 4.1.6 ECP
    - 4.1.7 Eye and Eyelid Movement
    - 4.1.8 Pupillary Dilation
    - 4.1.9 Muscle Tension, Tremor
    - 4.1.10 Heart Rate, Hear Rate Variability, Blood Pressure
    - 4.1.11 Breathing Analysis
    - 4.1.12 Body Fluid Analysis
    - 4.1.13 Handwriting Analysis
  - 4.2 Combined Physiological Measures
  - 4.3 Speech Pattern Analysis

TABLE 3

Applicability Matrix of Workload Methodologies Across Universal Operator Behaviors

WORKLOAD METHODOLOGIES		UNIVERSAL OPERATOR BEHAVIORS						
		1.1 Searching for and Receiving Information	1.2 Identifying Objects, Actions, and Events	2.1 Information Processing	2.2 Problem Solving and Decision Making	3. Communication Processes	4.1 Simple/Discrete Motor Processes	4.2 Complex/Continuous Motor Processes
1.1	Rating Scales	3	3	3	3	2	3	3
1.2	Interviews and Questionnaries	3	3	3	3	2	3	3
2.1.1	Task Component, Time Summation	3	3	3	3	3	3	3
2.1.2	Information-Theoretic	1	0	2	0	0	0	1
2.2.1	Nonadaptive, Arith./Logic	3	3	3	2	2	3	3
2.2.2	Nonadaptive, Tracking	3	3	2	2	2	2	2
2.2.3	Time Estimation	0	0	2	0	2	0	1
2.2.4	Adaptive, Arith./Logic	0	0	0	0	0	2	2
2.2.5	Adaptive, Tracking	1	0	1	0	0	2	2
2.3	Occlusion	1	1	1	1	0	1	0
3.1	Single Measure-Primary	1	1	1	1	1	1	1
3.2	Multiple Measure-Primary	1	1	1	2	0	1	2
3.3	Math. Modeling	0	0	1	1	0	1	1
4.1.1	FFF	1	1	1	1	1	0	0
4.1.2	GSR	1	1	1	1	0	1	1
4.1.3	EKG	1	1	1	1	0	2	2
4.1.4	EMG	2	1	2	2	1	2	2
4.1.5	EEC	0	0	1	0	0	0	0
4.1.6	ECP	1	1	1	1	1	1	1
4.1.7	Eye and Eyelid Movement	1	1	1	1	2	0	0
4.1.8	Pupillary Dilation	2	2	2	2	1	2	2
4.1.9	Muscle Tension, Tremor	1	1	1	1	1	2	2
4.1.10	Heart Rate, Heart Rate Variability, Blood Pressure	1	1	1	1	0	0	1
4.1.11	Breathing Analysis	1	1	1	1	1	1	1
4.1.12	Body Fluid Analysis	1	1	1	2	1	1	2
4.1.13	Handwriting Analysis	1	1	2	2	2	0	0
4.2	Combined Physiological Measure	1	1	1	1	1	0	1
4.3	Speech Pattern Analysis	0	0	1	1	1	0	0

<u>Weightings</u>

0 = No research support or only negative support
1 = Limited research support; some conflicting data
2 = Limited research support; no conflicting data
3 = Well documented research support

TABLE 4

Worksheet for Guide in Selecting a Workload Assessment Methodology for Aircrew Flight Test and Evaluation

UNIVERSAL OPERATOR BEHAVIORS

WORKLOAD METHODOLOGIES

BEHAVIOR CHECK (✓) OR WEIGHTING		1.1 Searching for and Receiving Information	1.2 Identifying Objects, Actions, and Events	2.1 Information Processing	2.2 Problem Solving and Decision Making	3. Communication Processes	4.1 Simple/Discrete Motor Processes	4.2 Complex/Continuous Motor Processes	SUM OF BEHAVIOR WEIGHTINGS FOR EACH METHOD
1.1	Rating Scales								
1.2	Interviews and Questionnaires								
2.1.1	Task Component, Time Summation								
2.1.2	Information-Theoretic								
2.2.2	Nonadaptive, Arith./Logic								
2.2.3	Time Estimation								
2.2.4	Adaptive, Arith./Logic								
2.2.5	Adaptive, Tracking								
2.3	Occlusion								
3.1	Single Measure-Primary								
3.2	Multiple Measure-Primary								
3.3	Math Modeling								
4.1.1	CFF								
4.1.2	GSR								
4.1.3	EKG								
4.1.4	EMG								
4.1.5	EEG								
4.1.6	ECP								
4.1.7	Eye and Eyelid Movement								
4.1.8	Pupillary Dilation								
4.1.9	Muscle Tension, Tremor								
4.1.10	Heart Rate, Heart Rate Variability, Blood Pressure								
4.1.11	Breathing Analysis								
4.1.12	Body Fluid Analysis								
4.1.13	Handwriting Analysis								
4.2	Combined Physiological Measure								
4.3	Speech Pattern Analysis								

TABLE 5

Worksheet for Guide in Selecting a Workload Assessment Methodology for Aircrew Flight Test and Evaluation SS-3/IRDS Addition Example

UNIVERSAL OPERATOR BEHAVIORS

WORKLOAD METHODOLOGIES		1.1 Searching for and Receiving Information	1.2 Identifying Objects, Actions, and Events	2.1 Information Processing	2.2 Problem Solving and Decision Making	3. Communication Processes	4.1 Simple/Discrete Motor Processes	4.1 Complex/Continuous Motor Processes	SUM OF BEHAVIOR WEIGHTINGS FOR EACH METHOD	RANKING
BEHAVIOR CHECK (✓) OR WEIGHTING		2	5	2	5	4	2	3		
1.1	Rating Scales	6	15	6	15	8	6	9	65	2
1.2	Interviews and Questionnaires	6	15	6	15	8	6	9	65	2
2.1.1	Task Component, Time Summation	6	15	6	15	12	6	9	69	1
2.1.2	Information-Theoretic	2	0	4	0	0	0	3	9	
2.2.1	Nonadaptive, Arith./Logic	6	15	6	10	8	6	9	60	4
2.2.2	Nonadaptive, Tracking	6	15	4	10	8	4	6	53	5
2.2.3	Time Estimation	0	0	4	0	8	0	3	15	
2.2.4	Adaptive, Arith./Logic	0	0	0	0	0	4	6	10	
2.2.5	Adaptive, Tracking	2	0	2	0	0	4	6	14	
2.3	Occlusion	2	5	2	5	0	2	0	16	
3.1	Single Measure-Primary	2	5	2	5	4	2	3	23	
3.2	Multiple Measure-Primary	2	5	2	10	0	2	6	27	
3.3	Math Modeling	0	0	2	5	0	2	3	12	
4.1.1	FFF	2	5	2	5	4	0	0	18	
4.1.2	GSR	2	5	2	5	0	2	3	19	
4.1.3	EKG	2	5	2	5	0	4	6	24	
4.1.4	EMG	4	5	4	10	4	4	6	37	
4.1.5	EEG	0	0	2	0	0	0	0	2	
4.1.6	ECP	2	5	2	5	4	2	3	23	
4.1.7	Eye and Eyelid Movement	2	5	2	5	8	0	0	22	
4.1.8	Pupillary Dilation	4	10	4	10	4	4	6	42	6
4.1.9	Muscle Tension, Tremor	2	5	2	5	4	4	6	28	
4.1.10	Heart Rate, Heart Rate Variability, Blood Pressure	2	5	2	5	0	0	3	17	
4.1.11	Brathing Analysis	2	5	2	5	4	2	3	23	
4.1.12	Body Fluid Analysis	2	5	2	10	4	2	6	31	
4.1.13	Handwriting Analysis	2	5	4	10	8	0	0	29	
4.2	Combined Physiological Measure	2	5	2	5	4	0	3	21	
4.3	Speech Pattern Analysis	0	0	2	5	4	0	0	11	

TABLE 6

Feasibility of Workload Techniques for In-Flight Environments

CRITICAL CRITERIA

WORKLOAD METHODOLOGIES		Physical Space Required	Portability	Intrusion-Safety	Data Transmission & Recording	Experimental Control	Integration into Aicraft System	Crew Acceptance
1.1	Rating Scales	S	S	S	S	P	S	S
1.2	Interviews and Questionnaires	S	S	S	S	P	S	S
2.1.1	Task Component, Time Summation	S	S	S	S	S	S	S
2.1.2	Information-Theoretic	S	S	S	S	S	P	S
2.2.1	Nonadaptive, Arith./Logic	S	S	S	S	P	S	S
2.2.2	Nonadaptive, Tracking	S	S	P	S	P	S	P
2.2.3	Time Estimation	S	S	S	S	P	S	S
2.2.4	Adaptive, Arith./Logic	P	S	S	P	S	P	S
2.2.5	Adaptive, Tracking	P	S	P	P	S	P	P
2.3	Occlusion	P	S	P	S	S	P	P
3.1	Single Measure-Primary	S	S	S	S	S	S	S
3.2	Multiple Measure-Primary	S	S	S	S	S	S	S
3.3	Math Modeling	S	S	S	S	S	S	S
4.1.1	F/F	P	S	P	S	S	P	P
4.1.2	GSR	S	S	S	S	S	S	P
4.1.3	EKG	S	S	S	S	S	S	P
4.1.4	EMG	S	S	S	S	S	S	P
4.1.5	EEG	S	S	S	S	S	S	P
4.1.6	ECP	S	S	S	S	S	S	P
4.1.7	Eye and Eyelid Movement	S	P	P	P	S	P	S
4.1.8	Pupillary Dilation	S	P	S	P	P	P	S
4.1.9	Muscle Tension, Tremor	S	P	S	S	S	S	S
4.1.10	Heart Rate, Heart Rate Variability, Blood Pressure	S	S	S	S	S	S	P
4.1.11	Breathing Analysis	S	S	S	S	S	S	P
4.1.12	Body Fluid Analysis	S	S	S	S	S	S	S
4.1.13	Handwriting Analysis	P	S	P	S	P	S	S
4.2	Combined Physiological Measure	S	S	S	S	S	S	P
4.3	Speech Pattern Analysis	S	S	S	S	S	S	S

Weightings

S: Solvable without difficulty; Problem does not exist.
P: Potential problem; Difficulty will be encountered.

WORKLOAD ASSESSMENT METHODOLOGY DEVELOPMENT

by

Billy M. Crawford
Systems Research Branch
Human Engineering Division
Wright-Patterson AFB OH 45433

The Workload Problem

During the development of advanced man-machine systems a number of important questions must be resolved. Many of them relate to human performance or manning requirements. For example: How much attention is required by operator tasks? Which tasks can be assigned to a single operator? How long can an operator perform his task effectively without a rest break? How much learning or training is necessary? What is the minimum crew size for a system? How will time pressure and other stresses affect task and ultimately mission performance? All the foregoing questions relate to performance and workload.

The designer/planner, based on his appraisal of the possible contingencies, typically attempts to minimize the frequency, extent and seriousness of work overload situations. However, he can neither personally nor vicariously, through others, rigorously assess the workload, or potential workload without a standard metric for adequately defining and quantifying it. Even if he does identify particular periods of potential workload excess, he does not, except in extreme and obvious cases, have quantitative information to assist in deciding which of the instances are the most critical and demanding and hence should, within resources and technological limitations, be given priority consideration in design. Nor does he have a criterion by which he can decide and demonstrate that the problem has been reasonably resolved.

In the development laboratories, alternative proposed designs or arrangements, or alternative procedures, may be compared on the basis of speed, accuracy, or errors. However operationally significant differences may not be revealed simply because the subjects are able to, and do, master their resources ("try harder") and thus compensate for what would otherwise be real differences.

Work overload at the mental or "cognitive" level has been associated with increases in the United States Air Force aircraft accident rate (Miholick, 1978). For example, during 1977 and 1978 "channelized attention" or "distraction" were factors in 16 accidents involving the loss of 12 aircraft, 9 fatalities, and a dollar loss of over 81 million dollars. "Task saturation" which results in intense concentration on the task perceived to be most important at the expense of other critical performance requirements was classified as "channelized attention." "Distraction" was used to refer to occasions in which an unexcepted task causes attention to be diverted to coping with the cause of the unscheduled task load.

If we are to make safe, economical use of human and material resources it is necessary to determine efficient crew compositions, appropriate assignments of duties and responsibilities to crew members, and effective allocations of functions and tasks among men, machines and computers (including software). In addition, it is necessary to identify the critical periods in a task or mission during which the operator's performance is particularly prone to degradation or failure because of work-overload stress. Further, it is necessary to provide improved, valid and quantitative methods for assessing equipment and system design, and procedural alternatives; and for mission planning and survivability/vulnerability analyses, to locate and quantitatively define the most critical and demanding task segments. In a parallel view, it is necessary to identify and quantitatively define those periods, if any, of sub-optimal workload stress so that the resources can be used elsewhere, or so that provisions can be made to preclude or alleviate boredom, loss of "sharpness" or alertness, etc., the effects of which can carry over to and jeopardize performance in subsequent tasks or mission periods. Due to the rapid advances in computer technology and the more centralized role computers assume in advanced systems, emphasis probably should be on man-computer interactions and information processing/decision-making functions which are not adequately accounted for by conventional human performance metrics, task analysis, time-and-motion, and time-line methods.

The principal objectives of a supportive workload research and development program should be (1) establishment of a set of theoretically-consistent component functions descriptive of the performance of crew members in relevant system tasks; (2) development of quantitative (mathematical) expressions of relationships between input-output parameters for the component functions and appropriate combinations thereof; (3) integration of the results of (1) and (2) above into a task analytic/computer modeling methodology; and (4) validation of the analytic/predictive methodology in a system design, development and test effort. Examples of approaches and metods contributing to achievement of the above objectives follow.

Adoption of a Workload Concept

Ryan (1947) addressed the problem of measuring the cost of sedentary, or "mental," work some 30 years ago in his text on the psychology of production. His concept of effort, presented in the same context, is similar to the concept of workload as it is used today. Ryan identified four possible meanings for effort:

(1) energy consumption,
(2) cost of work (e.g., fatigue, loss of health, dissatisfaction, etc.),
(3) aspects of psychological functioning which describe the "experience of the worker as he performs his job, and
(4) the rate of performance of an individual in relation to the maximum possible rate of performance under the given conditions.

Ryan indicated his preference for the latter (fourth) meaning of effort probably because it required that task performance be not only measured but also related to the capacity of the worker. The discussion of topics which ensues is based upon the assumption that effective resolution of workload problems depends upon the capability to measure, by a common metric, both task demands deriving from work situations and the inherent capabilities of the worker to meet them.

Performance Theory Development and Application

In order to progress in an orderly, systematic manner, it is necessary to explain and relate pertinent facts in a logically consistent manner. Current human performance theory can serve that function in a workload assessment program. The primary goal of human performance theory is to analyze human capabilities in a manner which will permit (1) identification and description of basic, component functions and (2) quantification of the limits of capacity in each component function. Theories which treat the human as principally an information processor of limited capacity appear to be most appropriate. Some of the research issues which have been associated with the development of such a theory are revealed by the following "types" of theories:

(1) Single Channel Theory (Welford, 1952; Broadbent, 1958). The human is strictly a "serial" processor.

(2) Undifferentiated Capacity Theory (Moray, 1967; Kahneman, 1973). The human behaves much like a time-sharing computer with task interference strictly a function of total demand rate rather than specific to the nature of the processing tasks competing for capacity.

(3) Limited Capacity Central Mechanism Theory (Posner and Keele, 1970). Some, but not all, processes require the "central mechanism"; hence, parallel, as opposed to serial (Single Channel), processing is sometimes possible.

An example of current theorizing based largely on the single channel concept is that of W. H. Teichner. For the past several years various U.S. Government agencies sponsored efforts of Teichner to develop a general theory of human performance. The goal was a systematic approach to prediction of human peformance as a function of task variables and environmental factors (Teichner and Olson, 1971). Teichner drew heavily upon the available experimental psychology and physiology literature to identify empirical relationships and develop models of simple tasks which could be combined into a more comprehensive model or theory or used in predicting performance in more complex tasks (Teichner, 1974).

Based on observations of people engaged in a large variety of work situations, Teichner concluded that the same general functions comprise the various human activities involved; hence, the feasibility of modelling any human activity in terms of a finite set of generic subtasks. Teichner and Olson held that tasks always involve a transfer of information from an initial input to a final output. In other words, the human is a system which functions through a series of communication links and subtasks and that system is the same whether flying an airplane or dialing a telephone. No matter how the man-machine system context varies, at a given level of human system analysis the only differences will be in the activity or degree of loading of the subtasks.

Teichner's theoretical approach is consistent with human engineering and system analysis tradition in referring to man and machine as components of man-machine systems. Any operation on information within a component, whether man or machine, is called a "process" whereas transfers of information between components are called "tasks."

Although both the maximum complexity and maximum capacity of the human are constant according to Teichnerian theory, system capacity may be varied in a number of ways. For example, since operations may be performed by different combinations of available generic subtasks, it may be possible to replace the limiting function in a serial process with a higher capacity subtask. Or, the system may be redesigned for parallel processing at the limiting stage by allocating the function to another component, e.g., a machine or another person. Assuming the human is a single channel system, the maximum processing rate can be no greater than the capacity of the lowest capacity stage in a sequence, of course.

In developing his performance theory, Teichner bypassed the task taxonomy problem and went directly to empirical relationships and principles which could be used to predict dependent measures. The theory builds upon Donders' Law which is based on data obtained from attempts to measure the physiological time of mental processes associated with discrimination and choice in 1868 (Woodworth and Schlosberg, 1955). Donders' Law simply states that choice reaction-time (CRT) is composed of simple reaction time (a constant) stimulus categorization time, and response selection time.

Teichner initially modified Donders' Law as follows: (1) Stimulus identification time was included in simple reaction-time; (2) Stimulus code-to-response code translation time (T_{s-r}) was substituted for the response selection component; (3) Stimulus code-to-stimulus code translation time (T_{s-s}) was added to account for tasks in which it was necessary to transform one stimulus code to another before selecting a response; and (4) another component (c) was added to cover time required to select the motor program for executing the response. The resulting equation was:

$$CRT = a + T_{s-s} + T_{s-r} + c$$

in which "a" includes both stimulus encoding time and neural transmission time.

Teichner adopted a response criterion model proposed by McGill (1963) and Grice (1968) in order to account for empirical evidence that the "a component" of the above equation depends upon stimulus intensity and duration (Teichner and Krebs, 1972).

Teichner proposed to use coding theory and information metrics to quantify S-S translation. Two examples of S-S translation are compression and classification. Compression is exemplified as follows: Assume a four message, binary source encoded thus: 0001, 0010, 0100, and 1000 with equal probabilities of occurrence. Compression could be achieved by recoding, e.g., 00, 01, 10, 11, with no change in message probability. The average value of the original, or source code (L_s), is 4 bits per message as compared to 2 bits per message for the recoded messages (L_c). In coding theory, the average compression for a sequence of symbols is called the compression coefficient and is represented by the equation: $m = L_c/L_s$. The value of stimulus compression is in its effect on the S-R translation process. Because there is less information in the compressed message, the S-S translation should involve less time and error. Obviously the loss resulting from the compression process must be less than the gain at the S-R stage for it to be worthwhile.

The second form of S-S translation identified by Teichner is classification which results in a reduction in the number of messages S-S classification is exemplified as follows: Assume a message set of four, e.g., F1, F2, B1 and B2. This message set may be sorted into F (fighter) and B (bomber), a case of four-to-two mapping.

It can be seen that the cost effectiveness of S-S translations as described above is assessible in terms of changes in the information transmission rate (R) achieved for CRT tasks. The cost effectiveness index for compression (CE_c) is $CE_c = R/m$. CE_c is the rate of information processing per unit of compression. The cost effectiveness of reduction in messages through classification (CE_r) is expressed as follows: $CE_r = R/H_c/H_s$ where H_s is the amount of information in the original message set and H_c is the amount of information in the set after classification.

The same cost effectiveness concepts may be applied to the S-R translation process, in which case the recoded message is a response and is defined by a response code. Again, the impact of reduction or compression is expected to be greater speed and accuracy of response selection.

Teichner clearly distinguishes between response selection and response execution. It is assumed that responses are always defined symbolically by response codes. Only after the appropriate response code has been matched with the stimulus code does the associated motor response begin. The ensuing response execution may entail a series of effector selections whether the response modality is limb movement, body movement, or speech. Execution time will depend on factors such as distance travelled, amount and direction of force exerted, etc.

Teichner's thinking toward a complete theory, or model, of performance is represented by the flow diagram in Figure 1. The "a" component of his equation for CRT derives from a combination of sensory register of his equation for CRT derives from a combination of sensory register and scanner functioning. The flow diagram shows that the response criterion applied by the scanner derives from long-term stimulus memory (LTM-S) which also establishes operating levels for activating systems and scanner rate. LTM-S also provides for selective tuning of the register so that thresholds of "energy cells" for expected stimuli will be lower than for unexpected stimuli. Sensory register properties are derived from Hubel and Wiesel (1962).

Teichner hypothesized that the human component obtains information transmission rates consistent with system demands by making speed-accuracy tradeoffs in the following manner. At the input stage, with experience at a task, an individual learns what stimuli to expect, how much stimulus evidence is required to respond, what sampling rate is required, and makes sensory register/scanner adjustments consistent with task demands relative to speed and accuracy. Depending upon the operations involved, a range of speed-accuracy variations may be available at the S-S and S-R translation stages. And, finally, at the output stage the response criterion may be adjusted upward or downward to favor either accuracy or speed depending upon the information transmission demands of the system.

Habituation is handled in a way consistent with Sokolov's (1963) neuronal model. When a novel stimulus passes to the S-S translation stage and cannot be matched with a relevant event in LTM-S, the responsive register cell is tuned toward increasingly high threshold levels on successive occasions. When a stimulus event is detected by the scanner mechanism, a corresponding unit of short term memory (STM) is activated for a duration of time (e.g., 30 seconds) during which comparison can be made with LTM-S in support of the S-S translation. Teichner suggests that several available models are consistent with the latter process (Norman, 1970; Saunders, Smith and Teichner, 1974).

The importance of Teichner's theorizing to workload assessment rests largely in its potential impact on task analysis. Traditional human engineering task analyses provide an overwhelming amount of detail almost totally unrelatable to available theoretical concepts and principles. Part of the difficulty is attributable to the fact that the conceptual frames of reference tend toward anatomical rather than functional task descriptions. Teichner's goal was to systematize the description of operator tasks and performance at a generic level consistent with both the environment/performance literature and the operational situation. An attempt to verify the applicability of a portion of Teichner's theory for a system simulation will be summarized in a later section.

Theory Testing via the Divided Attention Paradigm

Data, such as that obtained by W. E. Hick (1952), relating reaction time to the amount of information transmitted, and to the degree of stimulus-response compatibility (Garvey and Knowles, 1954), caused the idea that independent associative links exist between each stimulus and response to be replaced by the concept of a mediating limited capacity central mechanism (or system). The single channel interpretation of this system (Welford, 1952, and Broadbent, 1958) holds that a signal entering the system dominates the entire channel from the time it was selected until the response is initiated. Any other contending signals are either filtered out or held in store and gated into the channel after the response to the previous signal. Increase in response time for each unit of information transmitted provided measures of the processing demands a signal places on the limited capacity system.

However, additional research, principally task interference studies, suggested a need to modify the single channel concept. While the single channel, or serial processing, model requires that the time to perform two tasks simultaneously should equal the total of the times required to perform each task alone, sometimes it is found to be much less (Keele, 1967). This suggests that in such instances some components of the separate tasks may be processed in parallel and, hence, do not require exclusive use of a single channel mechanism. Attempts to account for this apparent parallel, rather than serial, processing led to the two alternate theories.

One of the alternatives is the general, undifferentiated capacity theory which holds that interferences between tasks occurs only when the total number of non-specific "processing units" is exceeded by the demand. That is, task interference is not specific to the peculiar nature of competing task components, or operations, involved, but simply reflects an "overdraw" on the available pool of capacity units. Moray (1967) modified this interpretation somewhat by hypothesizing a limited capacity processor, similar to a time-sharing computer, which allocates from its undifferentiated processing capacity amounts consistent with the demands of operations performed on the signal.

The second alternative to a single channel theory derives from the proposition that some, but not all, operations performed by the human information processing system are channeled through the limited capacity central mechanism (Posner and Keele, 1970). Thus, operations which do not require the mechanism may proceed in parallel without ever interfering. While it has been suggested that the limited capacity mechanism may be either a single channel or a parallel processing system which processes multiple signals with reduced efficiency (Kerr, 1973), it may be that there are several limited capacity mechanisms each of which is peculiar to a particular type of signal, sensory mode, or operation. It has been suggested that the amount of interference between operations depends upon overlap between factors such as verbal or spatial demands (Brooks, 1967; Allport, et al, 1972). Perhaps, after the fashion of Spearman's theory of intelligence, there are central mechanisms peculiar to each of several "specific factors" whereas operations of a "general" nature are processed in parallel. (Incidently, Teichner preferred a serial processing model and was confident that he could account for any apparent contradictions before his theoretical development was complete.)

Divided attention effects produced by requiring subjects to attempt two tasks simultaneously provide an excellent basis for evaluating hypotheses generated by any of the three variations of limited capacity theory. This fact has been recognized by several theorists. The result has been a proliferation of secondary tasks beyond the rather large number produced by engineering psychologists during the 1950's and 1960's. During the latter era, numerous researchers tailored secondary tasks for compatibility with primary tasks and used them to evaluate the efficiency of alternative procedural or man-machine interface designs. Although the results were valuable to the specific applications, they made few contributions to a basic understanding or quantification of human performance capabilities and limitations because of the lack of standard methods and metrics. There were obvious practical reasons for that deficiency which have been identified and discussed by Knowles (1963).

There is also some justification for using a variety of secondary tasks in exploring issues derived from the limited capacity mechanism theories. However, the Sternberg task and associated model of information processing stages hold a great deal of promise as a more or less standard approach to both theory testing and reserve capacity measurement (Steinberg, 1969). In addition to its power in theory testing and development, which has been demonstrated by the late George Briggs and his associates, primarily under sponsorship by the USAF Aerospace Medical Research Laboratory and the Office of Scientific Research (Briggs, et al, 1969, 1970, 1972), the task provides a method for assessing reserve capacity for a variety of workload situations. Although relatively simple and readily learned, the Sternberg task facilitates manipulation and control of three key functions in information processing/decision making tasks: (1) input, (2) central processing, and (3) output. Both input and output are readily quantified in information metrics--a common measure to biologists/neurophysiologists, behavioral scientists, communications and computer system engineers and, hence, potentially a boon to effective system engineering including associated man-machine tradeoffs and functions allocation. Moreover, the Sternberg task is amenable to variations in stimulus (e.g., visual, auditory, tactile) and response (manual, vocal) mode making it adaptable to a variety of dual task situations.

The Sternberg task is a choice-reaction task which facilitates manipulation of the loading at Stage 2 (Central Processing) while holding the requirements on the other stages constant. Stage 2 loading is varied by changing the number of "positive set" items (e.g., letters, digits, tones) the subject must maintain in memory. In performing the task, a subject listens, or watches, for a stimulus cue, or memory "probe," while maintaining a readiness to respond via a response device "yes" or "no" depending upon whether the cue "matches" or "does not match" an item stored in memory.

In applying the Sternberg task to the study of divided attention effects, the Sternberg task is first administered alone to obtain "baseline" data for 3 or more different "memory loads," e.g., 1, 2, 3 and 4 items. The resultant data (using correct responses only since incorrect responses are held to a "negligible" level) is used to plot reaction time (on the ordinate) against memory load (on the abscissa). A linear equation is fitted to this data plot to obtain a straight line with a particular slope and y axis intercept value. Thus, the intercept reflects time required for Stage 1 and Stage 3. The slope of the line reflects central processing time, i.e., Stage 2. Then, by requiring subjects to perform the same Sternberg task simultaneously with a second task, which is treated as the primary or priority task, one can acquire information relative to the nature and amount of workload imposed by the second task. For example, if the slope of the equation for the Sternberg data plot changes between the baseline and dual task conditions, the second task imposes significant demands on Stage 2 or central processing. If the intercept changes, the demands of the second task occur at Stages 1 and/or 3. The amount of change involved can be quantified in terms of the information metrics, bits and bits/sec., to obtain an indirect indication of workload associated with the task under study.

The utility of the Sternberg task is readily apparent from a review of the research program pursued by Briggs and his associates at Ohio State and New Mexico State Universities. Brigg's research centered around efforts to isolate divided-attention effects within one or more of the four possible stages of the Smith (1968) task paradigm: (1) encoding processes, which entail registering, sampling and preprocessing of stimulus information; (2) central processing (detailed analysis of sampled information for stimulus identification and definition); (3) response decoding, and (4) response control and execution. The essence of the research program results are, perhaps, summarized most concisely by tracing the progressive expansion of the function proposed by Sternberg (1969) for describing the relationship between choice-reaction time (CRT)[1] and the size of the positive memory set. (The reader should recall that the Sternberg technique requires that a subject first memorize a set of items of size M. Then, at a later time, an item, or "probe: is presented and the subject responds as to whether the item is, or is not, a member of the memorized, or "positive: set. The major dependent measure is the time (CRT) from presentation of the "probe" item until the response is executed.)

Sternberg expressed the function so:

$$RT = a + b(M)$$

Data collected by Swanson and Briggs (1969) showed a logarithmic relationship between CRT and memory load which led to the postulation that response time is a function of central processing uncertainty (H_c), a metric from information or communication theory (Shannon, 1949). Hence, Sternberg's expression was modified to read:

$$RT = a + b(H_c)$$

Subsequently, Swanson and Briggs (1969) demonstrated that the intercept constant (a) was linearly related to the amount of information transmitted (H_t), a communication theory metric of response accuracy; thus, the expression became:

$$RT = c + d(H_t) + b(H_c)$$

An experiment by Briggs and Blaha (1969) suggested that b could be expressed in terms of the number of displayed items to be classified (D) and the equation was modified again, thus:

$$RT = c + d(H_t) + e(H_c) + f(H_cD)$$

Briggs and Swanson (1970) next varied the response load (R) in an experiment. The results showed that it could be partialed out, thus quantifying still another component of performance and the expression was now:

$$RT = i + j(H_t) + h(R) + e(H_c) + f(H_cD)$$

By relating this resultant equation to the Smith information processing paradigm, Briggs (1972) made estimates of the time required for specific functions of the human information processing system. For example:

Preprocessing time:	180-280 msc
Stimulus sampling rate:	6.5 bits/sec
Recoding (Teichner's s-s translation):	25 bits/sec
Transfer from long-term memory to active memory:	39 bits/sec
Stimulus classification:	16 bits/sec
Response Decoding:	6 bits/sec

This is the type of quantitative information and generic classification scheme which is needed to permit the desired state-of-the-art advance in analytic/predictive methodology to effectively complement task and time line analyses during system design. Of course, a great deal of theory development and testing remains to be done.

In 1974, Briggs, Johnson and Shinar took a step toward integrating the Sternberg/Smith information processing paradigm with fundamental decision-making research by using a Bayesian decision expression to account for the sequence of decisions made by a subject in a classification task. Thus, the link has been established between simple choice behavior and more complex decision processes to suggest something of the potential for expanding and validating basic performance theory applicable to critical command-control-communication system design issues.

Workload Assessment as an Aid to Design

Questions concerning the impact of digital avionics for pilot workload have provided an opportunity for preliminary tests of both performance theory and the divided attention paradigm in an applied setting (Crawford, Pearson and Hoffman, 1978). The opportunity developed as follows:

The evolution of compact digital computers has made possible the development of digital avionics information systems. Such systems promise a number of advantages to both aircraft designers and users. For example, when interfaced with multipurpose cathode ray tube displays and multifunction switches,

---

[1] See Woodworth and Schlosberg (1955) for a review of Donders classic research on simple and disjunctive reaction time in 1868.

digital computation and storage capabilities can be used to reduce the number of dedicated instruments competing for cockpit panel area. Information which is not required by the pilot on a continuous or frequent basis can be stored and presented on demand either automatically, as related programmed mission events transpire, or in response to manual control actions (Zipoy, Premselaar, Gargett, Belyea and Hall, 1970). And with reduced demands for panel space, it will be easier to locate the multipurpose controls and displays in prime reach and viewing areas.

However, experienced pilots are troubled by the prospect of possible added activity--both mental and physical--required to gain access to information which is normally on dedicated instruments. Should the demand for such activities occur during peak operator workload, the impact on mission success might not be offset by the increased calculating power, speed, or accuracy afforded by the digital processor. Hence, a study was planned to evaluate the impact of multipurpose control/display tasks on the pilot's reserve capacity. Of partucular interest was the question as the whether or not the maintenance of knowledge of procedures associated with multifunction keyboard operation reduced the operator's reserve capacity for making choices or decisions such as might be required to handle contingency situations during a mission. Another purpose of this study was to investigate the compatibility of keyboard operations with continuous flight control tasks.

A computer-based simulator was used to present and score the task situations investigated (Brandt and Wartluft, 1975). Of the three different tasks involved, two, flight control and communications/IFF switching functions, represented actual tasks in aircraft systems. The third was a variation of the Sternberg task which served as a test to measure cognitive reserve capacity under various primary task conditions. All three tasks were implemented within a fixed-base cockpit simulator.

The front panel of the cockpit was equipped with three CRT-type displays. The center display was used to present information concerning basic flight parameters in a moving tape format. The cockpit also contained a throttle with afterburner switch (left side panel) and a center-mounted joystick control which were used, in combination with the displayed flight information, to "fly" various maneuvers. Printed computer outputs of simulator performance data included both mean absolute and root mean square error relative to specified control values based on "fly to" instructions for altitude, heading, bank angle, pitch, indicated airspeed, vertical velocity, angle-of-attack, and g-load.

Between the front instument panel and left side panel was a multifunction keyboard (MFK). This MFK, in combination with the CRT on the upper left of the front panel and a numerical entry keyboard, which was also located on the instrument panel (lower left), was used to simulate a multifunction interface with digital avionics subsystems. Subsystems, functions and states were displayed on the CRT to complement the feedback afforded by back-projected legends on the MFK push button faces.

The Sternberg task procedure used in this study was as follows: At the start of an experimental session, the experimenter read to the subject a set of 1, 2, 4 or 6 letters of the alphabet. The subject was asked to retain the set in memory during the succeeding block of trials. The four sets used were as follows: A, AH, AHJQ and AHJQSX. (Such sets are referred to as "positive sets.") During the block of trials the subject was presented (via a cassette tape player connected to his headset) a series of test stimuli or "probes" to which he was to make one of two responses: (1) "yes," the test stimulus matches the positive set, or (2) "no," it does not match, and, hence, is a member of a negative set. The negative set included the 9 letters, B, C, E, F, G, I, L, R and Y. Negative and positive stimuli occurred with equal probability (.5). Letters within the two sets also occurred with equal likelihood. The average inter-stimulus interval was 5.5 seconds and ranged from 3 to 7 seconds. "Yes" was indicated by the subject's pushing forward on a thumb switch on the joystick controller used for flight control; "no" was indicated by moving the thumb switch backward, i.e., toward the subject. Reaction times were scored automatically to the nearest millisecond. If a subject did not respond within 2 seconds the trial was scored "no response."

Central processing uncertainty (H_c) values for this study are: 1.00, 1.50, 2.00 and 2.31 bits for the 1-, 2-, 4-and 6-item memory sets respectively. Because there is always a 2-choice response, response uncertainty (H_r) = 1.0 bit in each instance (Attneave, 1959).

Four male subjects were used in the study. They were paid volunteer university students with an age range of 20-24 years. During the experiment a nominal cash incentive system was implemented to encourage performance. The amount of the incentive was based on the subject's relative standing in the group with respect to task performance criteria for each session. For dual task conditions the incentive value was weighted so as to emphasize priority for the flight control task when it was present. The incentive was weighted in favor of the MFK task when it was paired with the Sternberg task.

Prior to the experiment proper each subject was trained on all three tasks. Training sessions lasted two hours and were scheduled 2-4 times per week. Each subject was trained until task performance measures appeared to asymptote. Then each subject was tested under six different conditions: three single conditions and three dual task conditions: Flight control, MFK and Sternberg choice-reaction task, alone; and flight control plus MFK, flight control plus Sternberg task and MFK plus Sternberg task. When the Sternberg task was combined with MFK, it occurred only during periods when the subject was awaiting instruction for an MFK task of a given difficulty level. This was consistent with the interest in measuring cognitive loads associated with anticipation of MFK tasks rather than actual performance of them. The single task conditions preceded the dual task conditions for all subjects.

The four levels of MFK task difficulty investigated were quantified in terms of the number of bits of information transmitted via the keyboard in performing the tasks. The average value for each level was: I-7 bits; II-11 bits, III-17 bits and IV-26 bits.

The type of maneuver "flown" was the independent variable for the flight control task. Although seven maneuvers were flown, preliminary analyses showed that not all maneuvers were discriminable in terms of the weighted tracking error scores. Hence, the maneuvers were combined into two groups labelled "easy" and "difficult." "Easy" maneuvers included straight and level flight and level turns. "Difficult" maneuvers were climbing and diving turns. The error scores (X) were comprised as follows: $X = (0.01)\,\Delta$ altitude + $(0.1)\,\Delta$ airspeed for straight and level and stall, $X = (0.01)\,\Delta$ altitude + $(0.1)\,\Delta$ airspeed + $(1.1)\,\Delta$ g-load for straight and level turns, and $X = (0.005)\,\Delta$ vertical velocity + $(0.1)\,\Delta$ airspeed + $(1.0)\,\Delta$ g-load for turning dives and climbs. The delta values represent average error, i.e., deviation from the prescribed "fly to" value for the given flight parameter, per unit of time on the task. Altitude was measured in feet, airspeed in knots and vertical velocity in feet/minute. The flight parameter combinations and associated weights for each maneuver type were based on pilot opinion and research findings summarized in a separate report (Woodruff, 1972). MFK performance on multifunction keyboard tasks was measured in terms of task time and errors. The dependent measure for the Sternberg task was reaction time. Errors and failures to respond within two seconds were also recorded.

A simple analysis of variance (repeated-measures design) was applied to the scores for the flight control single task condition. The difference between easy and difficult conditions was statistically significant ($p < .05$). The mean and standard deviations for the easy condition were 1.09 and 0.17. Corresponding values for the difficult condition were 5.11 and 1.51.

The effect of MFK task difficulty proved significant statistically ($p < .001$). Mean task times (seconds) and standard deviations (in parentheses) for the four difficulty levels were: I-3.97 (0.32); II-5.95. (0.53), III-7.43 (0.68), IV-9.87 (0.83). The average rate of information transmission via the MFK system varied from 1.8 bits/sec. to 2.6 bits/sec. across the four levels of MFK task difficulty.

The method of least squares was used to fit a straight line to the Sternberg data. The result is reflected by the following regression equation for the single task, or baseline, condition:

$$RT = 549 + 118(H_c)$$

Although mean flight control error was greater when the flight control task was combined with MFK tasks, the differences were not statistically significant. Similarly, MFK task times increased under dual task conditions, but the differences were not statistically significant. Flight control error scores were virtually identical for flight control alone as compared to flight control with the Sternberg task. The Sternberg task had no statistically significant impact on MFK task time.

The method of least squares was used to fit linear equations to Sternberg response time data for each dual task condition. This permits comparison of intercept and slope values with those obtained for the Sternberg task baseline condition, for the purpose of localizing divided attention effects within the four stage information processing model.

Preliminary analysis showed no significant differences between levels of MFK task difficulty in terms of slopes and intercepts. Hence, a single regression equation was derived for the combined MFK levels. Equations for the resultant three dual task conditions are as follows:

Sternberg with MFK "Rehearsal" $RT = 617 + 118(H_c)$
Sternberg with Easy Flight Control $RT = 694 + 98(H_c)$
Sternberg with Difficult Flight Control $RT = 855 + 31(H_c)$

F-tests (Snedecor and Cockran, 1967) indicate that (1) slopes and intercepts for the flight control conditions differ significantly from those for the baseline condition, and (2) the intercept value varies significantly between the baseline and MFK implicit rehearsal condition.

Interpreted in the traditional manner, the preceding results indicate that the effect of MFK "impicit rehearsal" is in the input or output stage of information processing only. Following the empirical evidence and logic of Briggs, et al (1972), the effect is probably in the input stage. The difference in intercept values amounts to a 12% average increase in input-output time attributable to MFK "implicit rehearsal."

Active flight control, on the other hand, appears to impact both input and central processing as evidenced by differences form baseline in both intercept and slope values for the regression equation. Moreover, there is an increase in input-output time (28% and 55% for easy and difficult flight control, respectively) and an increase in central processing rate. The central processing rate for the baseline condition is 8.47 bits/sec. as compared to 10.20 bits/sec. and 32.26 bits/sec. for the easy and difficult flight control conditions respectively. This increase in central processing rate under the dual task condition is consistent, with results obtained by Lyons and Briggs (Briggs, et al. 1972). It was attributed to the subject's conducting fewer or less complete tests of the probe stimulus under the greater loading conditions. This apparent switch in mode of operation in the central processing stage may prove to be a valuable aid to identification of significant workload changes.

The observed variations in Sternberg task response accuracy suggested the appropriateness of further information analyses, i.e., calculation of the average amount of information transmitted (which would reflect <u>all</u> the data, including erroneous responses and no responses. These values for the baseline and two levels of each dual task condition are presented below.

AVERAGE INFORMATION TRANSMITTED (H_t) IN BITS FOR STERNBERG TASK

	H_c			
Condition	1.00	1.50	2.00	2.31
Baseline	.85	.85	.88	.41
Easy MFK	.86	.85	.94	.42
Difficult MFK	.84	.88	.86	.32
Easy Flight Control	.82	.77	.79	.27
Difficult Flight Control	.72	.72	.79	.26

These data clearly indicate that the 5-item memory set (H_c = 2.31 bits) produced an overload situation for every task condition.

Effective Uncertainty Reduction. Since perfect performance is represented by H_t = 1.00 bit in each instance, the above tabled values were taken to represent percentage of the information reduction task effectively accomplished by the subjects. An information reduction task is defined as one in which the amount of uncertainty associated with the response is less than that associated with the stimulus (Coombs, Daves and Tversky, 1970). Thus, using the measures of central processing time, a set of "effective uncertainty reduction rates" were derived and plotted graphically as shown in Figure 2. Note the consis- increase in efficiency as H_c goes from 1.00 to 2.00 bits with the overload effect at H_c = 2.31 for all conditions. Further study of Figure 1 suggests that cognitive reserve capacity is reduced by 20, 31, 45 and 54 percent by the four primary tasks (easy MFK "rehearsal," control), respectively.

With regard to the design issue addressed by the foregoing study, it appears that tasks imposed by multifunction switch concept places demands on the operator which may detract from the value of digital processing capabilities in avionics systems. The concept necessitates the concentration of uncertainty, normally distributed among the various dedicated instrument control/display interfaces, at a single interface. Hence, uncertainty which is normally removed via separate controls and displays for each subsystem/function has to be eliminated via keyboard actions on each occasion that the operator interacts with the multifunction system. Thus, while the digitally-based MFK system is relatively efficient in terms of action and information transmission rates, the tasks are generally more complex and take longer than corresponding ones for dedicated instruments.

The MFK flight control simulation and data appeared to provide a good opportunity for evaluating the practicality of general functions incorporated by Teichnerian performance theory. One of the more complex MFK task sequences was selected for that purpose. The task involved the transmission of 40 bits of information via 13 steps or key actions. Tiechner's theoretical components were then "mapped on" to the MFK task sequence. Then a second laboratory simulation was generated by using cards with symbols on them to model the same set of theoretical task components included in the MFK task sequence. The card-symbol simulation was used to generate a set of performance data using students at the University of New Mexico. Although this effort was only exploratory and has not been formerly documented, reasonably good agreement between task time means and variances was obtained for the two sequences. Mean task time for the card task was 8.3 seconds as compared to 8.7 for the MFK task.

As a further step toward integration of performance theory, part-task simulations of operator workloads and system performance, a computer programmed model of the 40-bit MFK task components was developed using Systems Analysis of Integrated Networks of Tasks. (SAINT will be discussed in more detail in a subsequent section.) Close agreement was obtained between empirical data from the cockpit simulator and SAINT modeling output. One hundred-sixty interactions of the computer model produced a mean task time of 8.8 seconds.

Real-time simulations of operational tasks, as described above, are an essential part of the theory development and testing process which must precede the achievement of an adequate analytic, descriptive and predictive data base to effectively support workload allocation in man-machine systems.

Physiological Correlates of Performance

Another line of research promising significant insights into the basis of human workload capabilities and limitations at the neurophysiological level as well as providing intermediate workload assessment aids involves the measurement of physiological correlates of performance. In 1934, Luckiesh and Moss, lighting experts, reported data on the relationship between heart rate and illumination level for a reading task. The data showed decrements in mean heart rate as a function of task duration; moreover, the lower the lighting level, the greater the decrement. Luckiesh and Moss, interpreted the finding as indicative of the greater amount of effort required under low light level conditions. However, M. E. Bitterman (1948) in reviewing the lighting research literature completely discredited this notion of Luckiesh and Moss in the following words: ".... everything we know about cardiovascular functioning would lead to quite the opposite conclusion, i.e., that heart rate is directly rather than inversely related to the cost of work. Heart rate is positively correlated with metabolic rate which we know to be a direct index of energy expenditure, and Hadley (1941) has found a positive correlation between heart rate and muscular tension which Dr. Luckiesh himself accepts as an index of exertion in visual work."

Whether Bitterman was correct or not in his criticism of Luckiesh and Moss, it is interesting to note that they might have had a basis for appeal in the research of a physiologist, Darrow, who took an apparently corroborative position in 1939--five years after Luckiesh and Moss published, but prior to Bitterman's review.

Darrow (1939) reported data to support his postulation that both noxious stimuli and mental activity involving "associative processes" are accompanied by cardiac acceleration in contrast to attention to sensory stimuli requiring "no extensive association of ideas" which is accompanied by cardiac deceleration.

Twenty-six years later, Lacey (1965), having reviewed a large number of related experimental findings, rephrased and expanded Darrow's postulation by suggesting that behavioral arousal, electrocardiacal arousal, and autonomic arousal are different forms of arousal and that the associated activation processes reflect the intended aim or goal of behavior as well as its intensive dimension. In elaborating, Lacey noted that an increasing number of psychophysiological experiments demonstrated that different stimulus situations reliably produce different patterns of somatic response. Listening to auditory stimuli, looking at pictures, tapping telegraph keys, warm and cold stimuli--each condition produces a different pattern of somatic responses (Davis, et al, 1955; Davis, 1957). To illustrate, reception of external stimuli, with no motor response required, produces a heart rate decrease concomitant with the more "typical" increase in other autonomic responses, e.g., palmar conductance (Lacey, 1959; Lacey, et al, 1963; Obrist, 1963).

Without going into a detailed review of evidence cited by Lacey with regard to underlying physiological mechanisms and the complex nature of relationships between the cardiac response and cortical activity, perhaps it will suffice for the purpose of this general discussion to use Lacey's findings as an indication of the potential value of physiological correlates of behavior as a relatively unobtrusive, objective technique for analyzing task performance at the cognitive level and obtaining guidance with respect to the stage, or stages, at which work overload occurs for any given individual, or group of individuals.

Lacey and associates began by presenting eight "stressor-situations" in different orders to three samples of subjects. The situations could be ordered along a continuum in that some required only attentive observations of the environment, e.g., looking at an intermittently presented light, while others involved increasingly greater amounts of internal cognitive functioning--retrieval of information from memory and problem solving activity, as in mental arithmetic. The results consistently showed that sensory intake was associated with cardiac deceleration and restraint of systolic blood pressure whereas tasks at the other end of the continuum (internalized cognitive processing) produced large increase in heart rate and blood pressure. On the other hand, respiratory rate and palmar conductance showed the non-specific or nondiscriminant, actuation pattern consistent with Cannon-based "arousal" or "activation theory." Thus, depressor-decelerative processes are associated with facilitation of environmental intake; pressor-accelerative processes with filtering out irrelevant stimuli which interfere with central cognitive functioning. This finding was supported by Obrist (1963) using a different sample of subjects and different stimulus situations. Confirmatory evidence was obtained from additional studies which showed (1) attention to visual and auditory stimuli to produce cardiac deceleration while respiratory rate increased, (2) "thinking" to produce cardiac acceleration, and (3) the more "analytic" the child, the greater the acceleration (Kagan and Rosman, 1964; Kagan and Lewis, 1965; Lewis, et al, 1965). Moreover, in reaction time experiments, Lacey has found that the greater the cardiac deceleration in anticipation of the stimulus, the faster the motor response.

In summary, Lacey concludes that different fractions of autonomic, electroencephalographic, and motor response are mediated separately by mechanisms which are clearly dissociable although they may be closely related. He suggests that the biological utility of the dissociation resides in the capability of the different fractions of response to influence cortical and subcortical functioning different, sometimes opposing, ways.

Kibler (1967), in an Aerospace Medical Research Laboratory study effort, sought to bridge the gap between applications and laboratory research on the different cardiac response-stimulus situation relationships by means of a vigilance experiment. The resultant data showed a positive relationship between the extent of stimulu-oriented cardiac deceleration and detection efficiency during a 1 1/2 hour vigil. The study was regarded as a significant step toward developing an independent measure of alertness during vigilance tasks. Subsequently, an unpublished pilot study by Crawford and Bachert, also of the Aerospace Medical Research Laboratory, showed a trend toward increased cardiac deceleration, and reduced sinus arrhythmia (the tendency of the normal heart rhythem toward irregularity), as a function of decreased signal-to-noise ratios, produced by adding clutter to a simulated airborne digitized radar return display.

In the laboratory, Kalsbeek (1971) has found significant reduction in sinus arrhythmia as a function of increases in the signal rate in a perceptual motor task. Kalsbeek (1968) also reported data indicative of reduced arrhythmia as a function of increased task demands in a flight control simulation.

Cardiac data obtained from Navy carrier pilots flying missions over Southeast Asia showed average heart rates to be substantially higher during launch and recovery than during bomb runs (Plattner, 1967). These results were interpreted to mean bombing was a less demanding task than take-off and launch, which was somewhat surprising to the researchers although not necessarily to all pilots. It is conjectured that analysis of the specific stimulus-situations involved in accordance with Lacey's theoretical position might have reversed the interpretation.

Some attempts to use cardiac response measurement, in combination with a battery of other physiological correlates of performance, have proven less than satisfactory. One possible explanation for difficulties recognized in at least one such attempt is the failure to differentiate between actual workload and performance, i.e., removal of flight instrument information produced a decrement in flight control performance, which was interpreted as an overload condition; but it also reduced the information load, which if effectively processed would have resulted in improved performance. Careful, accurate data collection and analysis is also essential to effective use of physiological data within theoretical contexts as posed by Lacey.

Nevertheless, the evidence with regard to cardiac response "situational-specificity," is judged to be sufficient to warrant further investigation of the measure under carefully controlled conditions employing the Sternberg task and information processing paradigm to assess relationships between increasing demands of the various "stages," stimulus input, central processing, etc., as well as the transformation processes (classification, conservation, condensation, creation, etc.) at different demand levels. The ultimate potential advantages to this program are at least two-fold: (1) Increased validity of the performance theory developed and (2) a relatively unobtrusive, objective workload assessment technique for use during actual system operations and during system simulations to precisely identify crew functions which require automated aiding via digital processing capabilities.

Evoked potential measurement appears to be another technique with reasonable promise for facilitating performance theory and workload assessment developments. Instrumentation for obtaining average evoked potentials involves the attachment of electrodes to appropriate areas of the scalp in the same manner as required to produce an EEG. The continuous electrical activity so obtained is conducted through an amplifier to an averaging computer. A stimulus may then be presented, simultaneously averaged by the computer. The resultant measure of the nonrandom activity is the average evoked response (Childers and Perry, 1969).

This response-averaging technique, which enhances the signal-to-noise ratio, also accurately identifies specific psychological variables with components of the EEG. A stimulus initiates a series of physiological processes related to both perception and preparation for an overt behavioral response. Analysis of the electrical activity between stimulus and response can provide useful information concerning factors such as the timing, process speed and anatomical location of physiological events associated with the psychological processes involved. Cognitive and motivational as well as stimulus and response variables may be included in the experimental situations achieved via this arrangement (Vaughan, 1968).

Theoretical issues related to the limiting central mechanism and serial vs. parallel processing appear to be most amenable to investigation via evoked potential methodology. The value of evoked potential measures as an aid to assessment of workload under operational or system simulation conditions is yet to be established. However, Weissman (1969) in promoting the use of average evoked potentials for assessing the level that the technique has no equivalent when it comes to minimizing interference with the subject. Hence, evoked potential measurement must be considered possibly as an unobtrusive method for workload assessment under flight test or operational conditions.

It has been suggested that a complete battery of psychophysiological instruments might include the measurement of heart rate, electrical activity of the brain, muscle activity, skin resistance, blood pressure, sinus arrhythmia, average evoked potentials, urinalysis, parotid fluid, pupillary response, metabolic rate, oxygen uptake and ventilatory rate (Gartner and Murphy, 1976).

However, because of the prevalent interest in cognitive or information processing/decision making activities, the EKG and EEG domains currently have the greatest appeal as primary sources of psychophysiological data and continued exploratory development.

Analytic/Predictive Methodology

The final thrust of a comprehensive workload assessment development effort must include the incorporation of the results of products of the thrust areas into analytic and predictive methods. First the performance theory and quantitative functional relationships between human input-output parameters will have to be reflected in task analytic procedures. The purpose of task and analysis is to provide the basic building blocks for subsequent human engineering analyses during system design and development. Task analysis entails the specification of tasks to be accomplished by human operators including the behavioral requirements of the tasks, kinds of discriminations to be made, decision making, motor responses, etc. From the task analysis estimates of error rates, time line projections and personnel aptitude and training requirements must be made.

Task analytic methodology as it exists today represents little more than the crude beginning made some 25 or 30 years ago. Critically needed research required to appropriately expand and validate esential behavioral information has not been forthcoming. Consequently, job analyses are expected to do more than they possibly can. Although analysts continue to break work into smaller elements to produce the expected documentation, it is largely a reductionistic effort without sufficient regard to the meaningfulness of the behavioral elements (Bryan and Regan, 1963).

It is suggested that emphasis should be upon implementation of system models (mathematical and computer simulation models) as analytic/predictive tools during system design. It has been said that the sign of maturity in systems analysis will be the development of useful models (Shapero and Bates, 1959). The SAINT methodology promises to be a useful vehicle in achieving the desired advance in systems analysis (Wortman, Seifert and Duket, 1975). (SAINT was referenced briefly in the earlier discussion of simulation which was primarily concerned with real-time, man-in-the-loop simulations).

SAINT consists of a symbol set for modeling systems and a computer program for analyzing the models. SAINT includes the conceptual framework for representing systems which include discrete task elements, continuous state variables and interactions between them. SAINT is not a model. It simply provides a framework within which any quantitatively expressed model, or models, may be described and exercised. And, since it was designed for addressing human performance, in particular, within system contexts, it is potentially an ideal vehicle for integrating generic behavioral functions such as are advocated within human performance theory. The resultant computer models of systems concepts could, then, readily evaluate the probability and source of system/task demands which exceed operator, or crew, workload handling capability.

In applying SAINT, systems are represented as graphical networks of task-activities with which one or more operators interact. Each task is described with respect to how its performance relates to other tasks within the system of interest. The graphical analysis is then input to the SAINT computer program for automated performance assessment. Using Monte Carlo techniques, the SAINT program permits simulation of probabilistic task performance and precedence relationships while collecting estimates of system performance at the same time. Capabilities are included for simulating continuous or discrete system state variables and their response to discrete control task execution and for dynamic modification of both operator and system characteristics as dictated by internal or external simulated "events" (Kuperman and Seifert, 1975). Thus, this computer modeling technique permits fast time evaluation of human engineering design alternatives, and other human factors, e.g., skill level, training and motivation, within system contexts. However, it is just as dependent on a valid scientific base as conventional task analytic methodology.

Preliminary attempts have been made to apply SAINT to current USAF system design problems. For example, a SAINT model of the cockpit simulator used to investigate multifunction switching and multipurpose displays for the Digital Avionics Information System Advanced Development program was developed (Kuperman and Seifert, 1975). Model networks were developed for both conventional dedicated avionics subsystem instruments and the multipurpose controls and displays. Exercise of the model provided estimates of performance within the limits of available empirical data. Conclusions of the investigators included: (1) The SAINT simulation techniques are readily applicable to predictive modeling of new concepts of man/machine interaction. (2) The techniques are appropriate to the study of the theories of human performance and to evaluation of experimental metrics for their implementation.

REFERENCES

Allport, D. A., B. Antonis and P. Reynolds, "On the Division of Attention: A Disproof of the Single Channel Hypothesis," Qrtly Journal Experimental Psychology, 24, 225-235, 1972.

Bitterman, M. E., "Lighting and Visual Efficiancy," Illuminating Engineering, 906-931, 1948.

Briggs, G. E., "The Additivity Principle in Choice Reaction Time: A Functionalist Approach to Mental Processes," In Topics in Learning and Performance, Academic Press, N. Y., 1972.

Briggs, G. E., A. M. Johnson and D. Shinar, "Central Processing Uncertainty as a Determinant of Choice Reaction Time," Memory and Cognition, Vol. 2, 417-425, 1974.

Briggs, G. E., G. L. Peters and R. P. Fisher, "On the Locus of the Divided Attention Effects," Perception and Psychophysics, Vol. II, No. 4, 315-320, 1972.

Briggs, G. E. and J. Blaha, "Memory Retrieval and Central Comparison Times in Information Processing," Journal of Experimental Psychology, Vol. 86, 296-308, 1970.

Briggs, G. E., R. P. Fisher, S. N. Greenberg, J. J. Lyons, G. L. Peters, and D. Shinar, "Multi-Task Time-Sharing Requirements," AMRL-TR-71-105, W-PAFB, Ohio, 1972.

Broadbent, D. E., Perception and Communication, Pergamon, London, 1958.

Brooks, L. R., "The Suppression of Visualization by Reading," Qrtly Journal of Experimental Psychology, Vol. 19, 289-299, 1967.

Bryan, G. L. and J. J. Regan, "Training System Design," In H. P. Van Cott and R. G. Kinkade (Eds.) Human Engineering Guide to Equipment Design, Chapter 13, Revised Edition, U.S. Government Printing Office, Washington DC, 1972.

Childers, D. G. and N. W. Perry, The Human Visual Evoked Response, Charles C. Thomas, Publisher, Springfield, Illinois, 1969.

Crawford, B. M., W. H. Pearson and M. Hoffman, "Multi-Function Switching and Flight Control Workload," AMRL-TR-78-19, In Proceedings of Sixth Symposium on Psychology in the Department of Defense, USAF Academy, April 1978.

Crawford, B. M., W. H. Pearson, and M. S. Hoffman, Multipurpose Digital Switching and Flight Control Workload, AMRL-TR-78-43, W-PAFB, Ohio, 1978.

Crossman, E.R.F.W., "Entropy and Choice Time: The Effect of Frequency Unbalance in Choice Responses," Qrtly. Journal of Experimental Psychology, Vol. 5, 41-51, 1953.

Darrow, C. W., "Electrical and Circulatory Responses to Brief Sensory and Ideational Stimuli," Journal of Experimental Psychology, Vol. 12, 267-300, 1929.

Davis, R. C., "Response Patterns," Transactions, N. Y. Academy of Science, Vol. 19, 731-739, 1957.

Davis, R. C., A. M. Buckwald, and R. W. Frankman, "Autonomic and Muscular Responses and their Relation to Simple Stimuli," Psychological Monographs, Vol. 69, 1-71, 1955.

Gartner, W. B. and M. R. Murphy, Pilot Workload and Fatigue: A Critical Survey of Concepts and Assessment Techniques, NASA TN D-8365, Ames Research Center, Moffett Field, Calif., November 1976.

Garvey, W. D. and W. B. Knowles, "Response Time Patterns Associated with Various Display-Control Relationships," Journal of Experimental Psychology, Vol. 47, No. 5, 315-322, 1954.

Gerathewohl, S. J., E. L. Brown, J. E. Burke, K. A. Kimball, and S. P. Stackhouse, "In-flight Measurement of Pilot Workload: A Panel Discussion," Aviation, Space, and Environmental Medicine, 49 (6), 810-822, 1978.

Grice, G. R., "Stimulus Intensity and Response Evocation," Psychological Review, Vol. 75, 359-373, 1968.

Hadley, J. M., "Some Relationships Between Electrical Signs of Central and Peripheral Activity: II. During Mental Work," Journal of Experimental Psychology, Vol 28, 53-62, 1941.

Hick, W. E., "On the Rate of Gain of Information," Qrtly. Journal of Experimental Psychology, Vol. 4, No. 11, 1952.

Hubel, D. H. and T. N. Wiesel, "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex, Journal of Physiology, Vol. 160, 106-154, 1962.

Kagan, J. and B. L. Rosman, "Cardiac and Respiratory Correlates of Attention and an Analytic Attitude," Journal of Experimental Child Psychology, Vol. 1, 50-63, 1964.

Kagan, J. and M. Lewis, "Studies of Attention in the Human Infant," Merrill-Palmer Qrtly., Vol. 11, 95-127, 1965.

Kahneman, D., Attention and Effort, Englewood Cliffs, N. Y., Prentice-Hall, 1973.

Kalsbeek, J.W.H., "Objective Measurement of Mental Workload Possible Applications to the Flying-Task," Proceedings of the 55th AGARD Conference, Problems of the Cockpit Environment, 1968.

Kalsbeek, J.W.H., "Sinus Arrhythmia and the Dual Task Method in Measuring Mental Load," In W. T. Singleton, J. G. Fox, and D. Whitfield (Eds.), Measurement of Man at Work, pp 101-114, Taylor and Francis, London, 1971.

Keele, S. W., "Compatibility and Time-Sharing in Serial Reaction Time," Journal of Experimental Psychology, 75, 529-539, 1967.

Kerr, Beth, "Processing Demands during Mental Operations," Memory and Cognition, Vol. 1, No. 4, 401-412, 1973.

Kibler, A. W., The Relationship Between Stimulus-Oriented Changes in Heart Rate and Detection Efficiency in a Vigilance Task, AMRL-TR-67-233, W-PAFB, Ohio, February 1968.

Klemmer, E. T. and P. F. Mueller, Jr., The Rate of Handling Information: Key Pressing Response to Light Patterns, USAF Human Factors Operations Research Laboratory, Report No. 34, 1953.

Knowles, W. B., "Operator Loading Tasks," Human Factors, 5, 155-161, 1963.

Kuperman, G. G. and D. J. Seifert, "Development of a Computer Simulation Model for Evaluating DAIS Display Concepts," Proceedings Human Factors Society 19th Annual Meeting, pp 347-353, Dallas, Texas, October 1975.

Lacey, J. I., "Psychophysiological Approaches to the Evaluation of Psychotherapeutic Process and Outcome," In E. A. Rubinstein and M. B. Parloff (Eds.), Research in Psychotherapy, Washington DC, American Psychological Association, 1959.

Lacey, J. I., J. Kagen, B. C. Lacey and H. A. Moss, "The Visceral Level: Situational Determinants and Behavioral Correlates of Autonomic Response Patterns," In P. H. Knapp (Ed.), Expressions of Emotions in Man, International Universities Press, 1963.

Lewis, M., J. Kagan, H. Campbell, and J. Kalafat, "The Cardiac Response as a Correlate of Attention in Infants," Child Development, 1965.

Luckiesh, M. and F. K. Moss, "The Effect of Visual Effort upon the Heart-Rate," Journal of General Psychology, Vol. 31, 131-139, 1935.

McCormick, E. J., Human Factors Engineering (2nd Ed.), McGraw-Hill, N. Y., 1964.

McGill, W. J., "Stochastic Latency Mechanism," in R. D. Luce, R. R. Bush, and E. Galander (Eds.), Handbook of Mathematical Psychology, Vol. 1, New York: Wiley, 1963.

Miholick, Major, "Safety Trends for Commanders Attention: First Things First," in TIG Brief10, AFRP 11-1, The Inspector General, 1978.

Moray, N., "Where is Capacity Limited? A Survey and a Model," Acta Psychological, 27, 84-92, 1967.

Norman, D. A. (Ed.), Models of Human Memory, New York: Academic Press, 1970.

Obrist, P. A., "Cardiovascular Differentiation of Sensory Stimuli," Psychosomatic Medicine, Vol. 25, 450-469, 1963.

Pew, R. W., "Human Information Processing Concepts for Systems Engineers," In R. E. Machol (Ed.), System Engineering Handbook, McGraw-Hill, N. Y., 1965.

Pierce, J. R. and J. E. Karlin, Bell System Technical Journal, Vol. 36, 497, 1957.

Plattner, C. M., "Heart Strain Greater in Landing on Carrier," In Aviation Week and Space Technology, March 13, 1967.

Posner, M. I. and S. W. Keele, "Time and Space as Measures of Mental Operations," Invited Address, Division 3, American Psychological Association, Sep. 1970.

Quastler, H. and V. J. Wulff, Control Systems Laboratory, Report No. 62, University of Illinois, 1955.

Ryan, T. A., Work and Effort: The Psychology of Production, The Ronald Press Co., N.Y., 1947.

Saunders, R. S., M. G. Smith, and W. H. Teichner, Models of Short-Term Memory: A Critical Review, TR 74-1, Department of Psychology, New Mexico State University, Las Cruces, New Mexico, February 1974.

Shapero, A. and C. Bates, Jr., A Method for Performing Human Engineering Analysis of Weapons Systems, WADC TR 59-784, WPAFB, Ohio, 1959.

Smith, E. E., "Choice-Reaction Time: An Analysis of the Major Theoretical Positions," Psychological Bulletin, 69, 77-110, 1968.

Sokolov, E. N., Perception and the Conditioned Reflex, Trans. S. W. Waydenfold, Oxford: Pergamon Press, 1963.

Sternberg, S. The Discovery of Processing Stages: Extension of Donder's Method, In W. G. Koster (Ed.), Attention and Performance II, Acta Psychological, 30, 276-315, 1969.

Sternberg, S., "High-Speed Scanning in Human Memory," Science, 153, 652-654, 1966.

Sternberg, S., Memory-Scanning: Mental Processes Revealed by Reaction Time Experiments, American Scientist, 57, 421-457, 1969.

Swanson, J. M. and G. E. Briggs, "Information Processing as a Function of Speed versus Accuracy," Journal of Experimental Psychology, Vol. 81, No. 2, 223-229, 1969.

Teichner, W. H., Human Performance Simulation, Annual Report, AFOSR Contract No. F44620-76-C-0013, New Mexico State University, 1976.

Teichner, W. H., Quantitative Models for predicting Human Visual/Perceptual/Motor Performance, Final Technical Report 74-3, Contract ONR N0014-70-A-0147-0002, New Mexico State University, Las Cruces, New Mexico, October 1974.

Teichner, W. H. and D. E. Olson, "A Preliminary Theory of the Effects of Task and Environmental Factors on Human Performance," Human Factors, Vol. 13, No. 4, August 1971.

Teichner, W. H. and M. J. Krebs, "Laws of Simple Visual Reaction Time," Psychological Review, Vol. 79, 344-358, 1972.

Teichner, W. H. and M. J. Krebs, "Laws of Visual Choice-Reaction Time," Psychological Review, Vol. 81, No. 1, 75-98, 1974.

Vaughan, H. G., Jr., "The Relationship of Brain Activity to Scalp Recordings of Event-Related Potentials," In E. Donchin and D. Lindsley (Eds.C, Average Evoked Potentials: Methods, Results, and Evaluations, NASA SP-191, Proceedings of Conference at San Francisco, Calif., September 1968, pp 45-94.

Weissman, N. W., Preface, In E. Donchin and D. Lindsley (Eds.) Average Evoked Potentials: Methods, Results, and Evaluations, NASA SP-191, Proceedings of Conference at San Francisco, Calif., Sept 1968.

Welford, A. T., "The Measurement of Sensory-Motor Performance: Survey and Appraisal of Twelve Years of Progress," Ergonomics, Vol. 3, 139-230, 1960.

Welford, A. T., "The Psychological Refractory Period and the Timing of High Speed Performance: A Review and a Theory," British Journal of Psychology, Vol. 34, 2-19, 1952.

Woodworth, R. S. and H. Schlosberg, Experimental Psychology, Henry Holt and Co., Inc., 1965.

Wortman, D. B., S. D. Duket, and D. J. Seifert, "SAINT Simulation of a Remotely Piloted Vehicle Drone Control Facility, Proceedings, Human Factors Society, 19th Annual Meeting, pp 342-346, Dallas, Texas, October 1975.

QUANTITATIVE MILITARY WORKLOAD ANALYSIS

by

Richard A. Albanese
USAF School of Aerospace Medicine
Brooks AFB, Texas, 78235, USA

INTRODUCTION

A central goal of a military workload analyst is to understand the determinants of mission success in a military setting. The emphasis is on the human determinants of mission success with particular consideration to how the human uses the system he is given to accomplish the mission at hand. In quantitative workload analysis the final goal in many instances is to provide various numerical measures of mission performance. For example, when examining a bombing mission, a workload analyst using mathematical models might attempt to estimate the probability that bombs would land on target. Specifically, he might attempt a statement such as: "The estimated circular error probable is 250 feet given the present workload conditions." Other measures he might estimate include summary statistics such as anticipated loss rates against specific enemy defensive configurations, and rates of overall success against enemy targets, and these summary statistics will be of particular interest below.

A workload analyst studies the system under consideration to determine its capabilities and, when appropriate, he designs system changes or modifications with a view to improving system performance. The main purpose of this paper is to suggest that the workload analyst attempt to evaluate his proposed design modifications within the framework of a quantitative or semi-quantitative cost/benefit tradeoff. This is particularly appropriate when the analyst has developed relevant metrics describing system performance both with and without the system modification.

A workload analyst can suggest a wide variety of system changes ranging from hardware modifications to changes in system operating procedures. Whatever the changes suggested, a workload study in the military setting can be represented by a cost/benefit table as shown in Figure 1.

In Figure 1, the basic or unmodified system has effectiveness e, vulnerability v, and cost per system c. Subsequent discussion will provide definitions of e and v. System modifications can improve the effectiveness of a system from the point of view of making the system more capable of inflicting losses on the enemy. However, improving the fighting effectiveness of the system can increase or decrease the system's vulnerability, just as decreasing a system's vulnerability can either decrease or increase the system's fighting effectiveness. In Figure 1, e_1, v_1, and c_1 are the effectiveness, vulnerability and system's cost of the basic military system with system modification #1. The symbols e_2, v_2, and c_2 are used in a similar manner for the system with modification #2.

Perhaps most readers will agree that composing the table in Figure 1 is a step forward, but, of course, still remaining is the question of how to use the assembled data for actual decision making. Should an investment of money be made, and if so could modification #1 or modification #2 be purchased, or should one simply recommend that more elements of the basic system be procured? Analytical scenario modeling can be a decision aid in this circumstance, and this will be described in the following section. The method or methods whereby a quantitative tradeoff table such as that shown in Figure 1 can be developed will be described briefly in the section of this paper following the next concerning analytical scenario modeling.

ANALYTICAL SCENARIO MODELING

In this section, system mission effectiveness, e, and system mission vulnerability, v, will be defined in the context of analytical scenario modeling. For the purpose of illustrating the usefulness of analytical scenario modeling, a simple example from among a class of combat models called Lanchester models, will be employed. This class of models was developed by Lanchester, an aeronautical engineer, in about 1914, and is extremely simple in conception and approach (1, 2).

Consider two opposing forces, Blue force versus Red force. The rate of attrition of the Blue force should be proportional to the number of Red systems available, that is

$$dB/dt = - r \times R \qquad \text{Eq. 1.}$$

where B is the number of Blue force systems or elements, R is the number of Red force elements, and r is the constant of proportionality which reflects Red's ability to reduce Blue force. Similarly, the equivalent differential equation for the attrition of the Red force is where b is a constant of proportionality

$$dR/dt = - b \times B \qquad \text{Eq. 2.}$$

which measures Blue force's ability to reduce the Red force. If the Blue force is identified as the analyst's side, the proportionality constant b can be identified with system effectiveness e, and, similarly, the proportionality constant r can be identified with Blue force system vulnerability v. Thus the following equations obtain.

$$dB/dt = - v \times R \qquad \text{Eq. 3.}$$

and

$$dR/dt = - e \times B \qquad \text{Eq. 4.}$$

and these equations provide quantitative definitions of effectiveness and vulnerability. These equations are easily solved to provide B and R as functions of time t. These solutions are shown below for the interesting but highly simplified case where e and v are constants.

$$B(t) = (1/2\sqrt{e})\ \{(\sqrt{e}\ B_o + \sqrt{v}\ R_o)\exp(-\sqrt{ev}\ t) + (\sqrt{e}\ B_o - \sqrt{v}\ R_o)\exp(+\sqrt{ev}\ t)\} \qquad \text{Eq. 5.}$$

$$R(5) = (1/2\sqrt{v})\ \{(\sqrt{e}\ B_o + \sqrt{v}\ R_o)\exp(-\sqrt{ev}\ t) + (\sqrt{v}\ R_o - \sqrt{e}\ B_o)\exp(+\sqrt{ev}\ t)\} \qquad \text{Eq. 6.}$$

In these equations, B_o and R_o are the sizes of the Blue force and the Red force, respectively, at time $t = 0$ at the start of combat (prior to any losses). These last two equations describe force attrition during a battle. Far more complex attrition models are often developed to study force adequacy and tactics. The suggestion made here is that such attrition models or analytical scenario models be adapted and employed in workload tradeoff analyses. The concept is to compare design alternatives against predicted combat outcomes, and to choose that system modification which optimizes desired outcomes. This concept will be illustrated in the following by using equations #5 and #6.

A natural military goal is to reduce the enemy while minimizing one's own losses. This military goal can serve as an outcome metric which can discriminate between differing system modifications. Other outcome metrics can be defined such as minimizing one's own losses while reducing the enemy in the shortest time possible. However, for the purposes of the present illustration the simpler metric of minimizing losses alone will be employed.

Examining equations #5 and #6, it can be noted immediately that Blue unit will ultimately dominate the Red unit if the quantity $\sqrt{e}\ B_o$ is greater than the quantity $\sqrt{v}\ R_o$ (since $(\sqrt{v}\ R_o - \sqrt{e}\ B_o)$ is then a negative quantity in equation #6). If $\sqrt{e}\ B_o$ is greater than $\sqrt{v}\ R_o$, R will be zero at time $t = t^*_R$ where

$$t^*_R = (1/2\sqrt{ev})\ln((\sqrt{e}\ B_o + \sqrt{v}\ Ro)/(\sqrt{e}\ B_o - \sqrt{v}\ R_o)) \qquad \text{Eq. 7.}$$

and, using this critical time, maximum Blue force losses can be calculated using the following formula:

$$\text{Losses}_b = B_o - ((eB_o^2 - vR_o^2)/e)^{1/2} \qquad \text{Eq. 8.}$$

Similarly, if $\sqrt{v}\ R_o$ is greater than $\sqrt{e}\ B_o$ Red force will ultimately dominate Blue force, and B will be zero at time $t = t^*_B$ having lost all B_o systems.

These last equations will now be employed to accomplish an example tradeoff analysis. A hypothetical cost/benefit table is shown in Figure 2. In this table, the fact that $\sqrt{e_o}B = \sqrt{4} \times 250 = 500$ is less than $\sqrt{v}\ R_o = \sqrt{2} \times 400 - 565$, certainly motivates the Blue analyst to recommend changes. Modification #1 allows Blue to defeat Red while sustaining loss of 177 Blue elements. The cost of modification #1 is 37.50 million dollars which is a sum which would allow procurement of 37 additional unmodified systems. Since $\sqrt{4} \times 287$ is greater than $\sqrt{2} \times 400$, the Blue force augmented by 37 elements, would defeat the Red force, but in so doing the Blue force would sustain a loss of 201 systems. Thus, modification #1 would be preferred over the equivalent Blue force augmentation.

Now consider modification #2. This modification allows Blue force to win with a loss of 142 units. The cost of modification #2 is 125 million dollars. With this money, 125 additional unmodified systems can be procured to form a force of 375 fighting elements. With this size force, Blue force defeats Red force while sustaining losses of only 92 systems indicating that equivalent augmentation of the unmodified force would be preferable to purchasing modification #2. More complex mathematical models would allow consideration of purchases of various combinations of modifications #1 and #2. Putting these more complicated situations aside, and simply using what has been computed above, it can be concluded that if 125 million dollars were available, augmentation of the basic force should be accomplished without modifying the individual elements of the force. However, if 37.5 million dollars are available for use, modification #1 will minimize losses.

It has thus been illustrated how analytical scenario models can be used in workload analysis tradeoff studies. These models can help workload analysts define their earned return on investment and can help with decisions concerning modification alternatives. Admittedly perhaps the simplest scenario model has been employed here to illustrate scenario model usefulness. It is anticipated that real-world decisions would employ simulations which are far more complex and extremely well tested. Nonetheless, from the above very simple example, the workload analyst should be prepared to realize that in some instances it may be preferable to procure more of an unmodified system than to proceed to a modified system.

CONSTRUCTION OF QUANTITATIVE TRADEOFF TABLES

In this section, the construction of quantitative tradeoff tables as shown in Figures #1 and #2, will be briefly discussed. These tables can be constructed using three different types of data sets. A data set of type #1 consists of data derived on the military systems of interest and including the precise effectiveness and vulnerability figures needed to complete the tradeoff table. Type #1 data sets are rarely encountered in practice. These data sets can be developed from records of actual combat or they can be developed from records of realistic practice or training encounters where different systems are employed or compared. This data type, when it is available, provides the best and most direct data for tradeoff studies.

A data set of type #2 consists of data derived from the actual military systems of interest in the tradeoff study; however, the performance measures available from these systems are not the desired effectiveness and vulnerability measures. Often in this setting the available data are indirect measures of mission performance, or measures of human operator workload stress during mission performance, from which the likelihood of mission failure can be inferred. For example, when the concern is with a bombing mission, instead of obtaining the numbers of enemy targets destroyed per unit time by the competing systems, this data type might provide circular error probable figures from which enemy target destruction would have to be inferred. Still more indirect data concerning the bomber performance would be data that related to aircrew stress during the performance of trial missions. Such measures are, for example, voice stress measurements, galvanic skin responses, cortisol secretion and the like. It is clear that with data sets of type #2, the workload analyst faces a problem of extrapolating from the available performance measures to measures that are more relevant to a military decision in a tradeoff setting.

A data set of type #3 consists of data derived from systems which are not the military systems of interest or concern in the tradeoff deliberations, but are other military systems currently in the inventory, or are, as is often the case, laboratory simulations of the real systems under consideration. Thus data sets of type #3 also pose serious extrapolation problems. In this case, the extrapolation problem is one of relating data from one system to the relevant performance measure applicable to another system.

As discussed above, data sets of type #3 and type #3 require that the workload analyst extrapolate between measures, or between military systems, or both. [illegible] extrapolation can be done via experimentation or via the use of experimentation coupled with the application of mathematical models. The use of mathematical models in military workload analysis has been outlined in a previous publication wherein a coarse classification of available modeling techniques is provided.

SUMMARY

The above discussion has suggested that military workload analyses proceed in the setting of quantitative or semi-quantitative tradeoff analysis. This setting is already quite familiar to the hardware engineer, but may be a novel setting for the human factors workload specialist. The term semi-quantitative analysis is employed to recognize the fact that it will not always be possible to precisely quantitate effectiveness and vulnerability as well as one would wish.

The methods described in this report rely heavily on mathematical modeling techniques. This is seen in the suggestion to employ analytical scenario modeling in the tradeoff study, and is also seen in the suggestion to employ mathematical models in the construction of the tradeoff table from data sets that are not directly applicable. While mathematical models can be extremely useful and cost effective in application, they must be used with sober caution. Mathematical models are best employed with an attitude which considers the mathematical models, not as a replacement for traditional methods, but as an adjunct to commonly employed methods of analysis and deliberation. Mathematical models should in no way displace the direct use of experience and the direct consideration of empirical data. Rather, mathematical models should be used to enhance and highlight the utility of available data sources. The analyst's dictum "never believe your mathematical model" is a wise rule which is simply a statement of caution intending to remind the analyst that mathematical models are as fallible as any other human-contrived decision aid.

CONCLUSION

This report has discussed a method of tradeoff analysis as applied to workload analysis in the military environment. It is suggested that workload studies be performed in a tradeoff setting which allows the analyst to estimate the return on investment he has earned through his proposed system modifications. The methodologies described employ mathematical modeling techniques, and it is reinforced that these techniques are an adjunct to, and not a replacement of, more traditional methods of workload analysis.

REFERENCES

1. Lanchester, Frederick William. Mathematics in Warfare. IN: The World of Mathematics, Vol. 4, Simon and Schuster, New York, pp. 2138-2157.

2. Perla, Peter P. Approximation Techniques and Optimal Decision Making for Stochastic Lanchester Models. Technical Report #137. Department of Statistics, Carnegie-Mellon University, January 1978.

3. Albanese, Richard A. Mathematical Analysis and Computer Simulation in Military Mission Workload Assessment. AGARD CP-216, 1978.

VISUAL PERFORMANCE: A METHOD TO ASSESS WORKLOAD IN THE FLIGHT ENVIRONMENT

by

R. Simmons, M. Sanders, and K. Kimball
United States Army Aeromedical Research Laboratory
Fort Rucker, Alabama, 36362, USA

INTRODUCTION

Operator workload for the task of vehicle manipulation perhaps could be defined as the sum of sensory inputs, psychomotor responses, and cognitive processes. Sensory inputs to the operator are utilized to direct control manipulation, obtain feedback as to degree of effectiveness of the control movements, and to monitor system status. This input workload is combined with the psychomotor workload required to move the vehicle controls as dictated from the sensory inputs and feedback modes. More simply stated, workload measurements can be derived by objectively measuring the input and/or output of the operator.

The ability to manipulate an aircraft, as well as a tank, car, or any other vehicle, is directly related to inputs or cues the operator receives from the environment. Of these perceptual inputs (tactile, visual, auditory, etc.) required to fly an aircraft, visual cues are considered vital. E. Hartman has even estimated that vehicle operators acquire over 90% of their required information visually. Processing and integrating these visual cues allow the pilot to detect the aircraft's relative stability, ground references, and provide feedback from his control functions. During flight conducted under instrument meteorological conditions (IMC), lack of cues from the environment outside the aircraft requires the pilot to obtain the necessary visual information from instrument displays. As a consequence, there exists the need, independent of visual conditions, to determine what cues or visual workload are required to achieve maximum pilot efficiency with miminal fatigue-induced errors and safe mission accomplishment.

A great variety of apparatus and techniques have been developed for the study of visual performance/workload (2, 3, 4). One of the earlier devices was a smoked-drum Kymograph attached to the sclera of the eyeball via fine wire and barbed hooks. During the 1930's, electrooculography (EOG) techniques were developed which utilized electrodes placed around the eyes of the facial structure to monitor differential voltages as the eyeball was rotated (5).

The earliest documented technique for measuring the vital performance of pilots was to simply record pictures of the operator's face while he scanned the instruments (6). Improvements of this method were accomplished by arranging mirrors on the instrument panel and photographing the total arrangement. Documentation of eye movement was obtained by means of a camera mounted behind the pilot. During analysis a photo interpreter scanned the film to determine which mirror reflected the eye of the pilot at various times during the flight (7).

This technique was further refined by Mackworth (8). His approach was to mount a lightweight moving picture camera beside the operator's head along with a series of mirrors which reflected a dot representing the eye's motion. This dot was superimposed on photographs of the scene directly in front of center line of the head. More recently this same "corneal reflection" technique has been utilized by the US Army Aeromedical Research Laboratory in the study of Army pilot visual performance during helicopter flight (9, 10).

The corneal reflection technique is possible because of the smooth spherical front surface of the cornea. An incident beam of light can be partially reflected forming a bright spot or "highlight" on the cornea. The angle of the reflected light depends upon the angle between the incident light ray and a plane tangent to the reflecting surface. Since the cornea forms an eccentric bulge on the nearly spherical eyeball, the angle of this tangential plane on the cornea at any one point changes as the eye rotates around its center during eye movement. As a result, the position of the highlight follows the direction of movement of the cornea. The reflected beam is easily photographed on film. By mounting a camera lens on subject's head slightly above and between his eyes, the subject's normal visual field can be recorded and the highlight can be superimposed on the scene to give a constant eye reference to the eye's highlight, the area of visual concentration and the percentage of time for eye stabilization during any flight maneuver can be recorded.

Past research has demonstrated two major advantages of the corneal reflection technique for studying eye movement. First, the method is convenient for large scale testing of subjects in that it requires minimal training. Second, these studies have reported no significant interferance with normal eye movement (11, 12). This laboratory utilizes motion picture film to record the visual performance data. Figure 1 is a picture of the oculomotor lense and peripheral equipment. The total methodology is outlined in USAARL Report No. 77-4 (13).

Investigations which have been devised to collect data related to visual performance can be divided into three categories: (1) subjective opinions of visual performance, (2) objective visual performance data during fixed wing flight, and (3) objective data during helicopter flight. Studies by Siegel and MacPherson (14), Clark and Intano (15), Simmons, et al. (16) have analyzed the opinions of aviators as to which instruments they felt were utilized to fly selected maneuvers. However, these findings do not agree with the research results of Frezell, et al. (10), Sanders (12), and Simmons, et al. (13). These investigators have reported a very poor agreement between subjective data and actual pilot visual performance. Additional studies by Milton, Jones, and Fitts (6), Fitts, et al. (7), and Diamond (17) have utilized equipment to obtain objective visual performance data of aviators during flight maneuvers in several fixed wing aircraft. Although these investigations provided useful information as to visual performance during fixed wing flight, data obtained during this work cannot easily be generalized to rotary wing flight because of the extreme aerodynamic differences between airplanes and helicopters.

Sunkes, et al. (18), Sterm and Bynum (19), Frezell, et al. (9, 10) have recorded visual performance in helicopters during selected visual flight rules (VFR) flights. Additionally, two reports (20, 21) investigated a number of maneuvers utilizing both the interview technique as well as inflight recordings of visual performance of two aviators under instrument flight rules (IFR) conditions. These efforts have provided some needed information as to the frequency, duration, and sequence of fixations during helicopter operations.

Although these studies have provided useful information for the visual performance data base, much investigation remains to be accomplished before a reliable visual performance/workload model can be established for safe helicopter flight. The pupose of this report is to attempt to combine the visual performance investigations being performed at the US Army Aeromedical Research Laboratory into one mode for predicting visual workload.

THEORY

Several measurements of visual performance derived from data collected via the corneal reflection technique contribute to the total relationship of visual workload. In simple terms, oculomotor activity can be divided into two categories: (1) movement of the eye during which minimal information gathering occurs and, (2) fixation, a period of relatively no movement during which information transfer is felt to be the greatest (1). The movement activity is defined as the visual link value or the visual path traveled from one area of interest to another. On the other hand, the visual nonmovement term, visual fixation, is defined as stationary eye movement within a designated area for at least 100 milliseconds. Other visual terms which could be included are the total number of areas that are concentrated on (or fixated), the length of time of each fixation (or dwell time), and the frequency that areas of interest are fixated.

If one assumes that the major input mode is the fixation period, two possibilities exist. Visual workload could be a function of the time required for information to be transferred during fixation; or, workload could be related to the frequency of visits to an area of interest. Since from a previous investigation (22) neither term was found to adaquately describe visual activity independently, both comprise this input mode workload and should be combined. Thus, a formula utilizing these two terms would reflect the workload cost of all areas that were fixated by an operator during vehicle manipulation.

This formula would appear as: $CF_a = (T/\Sigma T+N/\Sigma N)/2$. CF_a represents the "cost factor" of an area of interest. "T" is lapse time spent fixated on the area divided by total time (ΣT) while "N" is the frequency of fixations of the area divided by the total number of fixations (ΣN). If these two values are divided by 2, the CF is in percentage of workload. If the CF values of several areas of interest lend themselves to being combined into common zones of interest, the CF values are simply summed together ($CF_a + CF_{a1} + CF_{a2} + CF_{a3} +...+...+$).

Based on our experience, the visual inputs required to manipulate an aircraft can be divided into three broad categories: (1) basic vehicle control, (2) barrier avoidance, and (3) navigational tracking. The first requirement takes precedence over the latter two. Under this category of basic vehicle control, visual workload can be further separated into three major zones of common areas visual interest. Again, the highest priority zone contains visual cues which provide information relating the basic vehicle stability about its three major axes of pitch, yaw, and roll.

The second zone of common areas of visual interest include the input information which supports the first areas but provides for more precise vehicle control. Information such as vehicle speed, altitude, and rates of acceleration would be provided from this zone.

The last zone would be comprised of vehicle status information. These cues would provide operator visual feedback as to the operational condition of the vehicle. Examples of such types of information would be provided from engine oil temperature, fuel pressure, or electrical gauges. As long as there were no malfunction of the vehicle as annunciated by one of these instruments, this zone of visual inputs would have the lowest priority of being monitored.

To summarize, the CF theory provides a method of combining numerous blocks of visual data to provide a more concise picture of input workload of vehicle operators. The CF value computed for Zone 1 should be an indicator of basic workload required to perform the task successfully. Zone 2 will also provide supportive data of the basic workload based on time available after Zone 1 requirements are met.

It should, however, be quickly pointed out that maximum visual performance of an area or zone of areas could indicate high visual workload. On the other hand, this same performance could reflect a high percentage of nonworkload (free time) in which the particular zone was fixated because it was centrally located. This could be demonstrated by similar visual workload in the central viewing field of a boat operator on a large lake and a helicopter pilot during nap-of-the-earth maneuvers. However, by establishing conditions which provide measurements of the baseline for the maximum time utilized and the minimum time required to maintain vehicle stability, these "free time" periods can be estimated. An example of this can be reviewed in USAARL Report No. 78-6, Visual Performance/Workload of Helicopter Pilots During Instrument Flight (22).

The remainder of this report all deal with the data base which the US Army Aeromedical Research Laboratory has obtained during helicopter and fixed wing maneuvers in attempting to establish a visual workload model. These types of data not only provide the needed information to test the CF theory, but also will provide information to improve and refine the theory to provide operational answers for safer military airborne operations.

APPLICATION

Initially, a study was designed to investigate the visual performance of helicopter pilots during actual flights under instrument flight conditions (IFR) (22). This study was unique because the aviators were forced by the test conditions to receive any and all visual cues to manipulate the aircraft from the instrument panel. This limited visual field allowed investigators to analyze which cues were fixated and derive what information was required by the pilots. During VFR this extraction of visual performance data would be very difficult because of the lack of precise definitions as to the quality of possible VFR cues.

Visual performance via the corneal reflection technique was collected from two groups of subject pilots. Subject groups were categorized on the basis of flight experience, with one group having over 2,000 more flight hours than the other. All subjects flew the same instrument flight profile comprised of eight basic maneuvers. The results of the study are summarized by Figure 2. IQS identifies the pilots with the most flight experienced while SQA represents the low time pilots. Z_1, Z_2 and Z_3 designate the three zones of instruments following the previously discussed method of classification. Table 1 is the listing of those instruments comprising each zone.

Since Zone 1 is the most critical indicator of visual workload, the data reflect that the experienced pilots had more workload to complete the mission than did the less experienced pilots. This could be further interpreted to mean that: (1) the IQA could spend less time in this flight environment before becoming fatigued, or (2) the IQA would most likely make more flight errors sooner than the SQA pilots. These results appear to contradict the common philosophy that experienced pilots should have been the better combat-prepared pilots. Therefore, the data were re-examined more closely for other possible explanations. In attempting to establish other group differences, it was concluded that although the IQA group did have the most total flight time, they were all currently holding job positions as instrument flight instructors. For this reason they, in fact, had less current "hands-on" experience than the SQA group who were all recent graduates of flight school and therefore had just completed a very concentrated block of "hands-on" flight experience.

To further test this line of thought, a single subject was selected who currently had 2,500 hours of flight experience but who had not flown for the past three years (23). His initial flight test results (NQA) are reflected by Figure 3. The results indicate a significantly higher level for his visual workload in Zone 1 to perform the same mission as the previous subjects. This subject was then given 14 hours of refresher training by the laboratory's instructor pilot. Figure 4 presents the results of his last flight (NQA) on the same profile compared again to the initial SQA subject group. It is apparent that his workload to perform the mission has been reduced to a similar level as that of the SQA group. These results would seem to indicate that utilizing the CF method of calculating visual workload aided in identifying differing visual workload as a function of aviator's current proficiency.

This same method was again assessed during a second investigation which compared the visual workload associated with flight of a fixed wing aircraft, during instrument conditions, compared to the original rotary wing instrument flights (24). AU-21 fixed wing aircraft was flown over the same flight profile as in the helicopter instrument flight. Two subject groups were again utilized. However, for this investigation, the first group were current instructor pilots (ICA) which compared to the IQA group in the helicopter report. The second group consisted of noncurrent U-21 pilots (NCA) who had not flown the U-21 for at least 3 years prior to the test flight. The purpose of this investigation was twofold in that it allowed a comparison of visual workload as a function of vehicle stability (i.e., rotary wing versus fixed wing aircraft) while further testing the currency versus experience question.

Figure 5 represents comparison of the two U-21 subject groups. The results indicate, as have past findings, that the noncurrent aviators (NCA) experienced more visual workload than did the current aviators (ICA). However, a confounding variable was that the NCA subjects were all current in the UH-1 helicopters. Because of this variable, the level of difference between subject groups is perhaps not as significant as would be anticipated. Nevertheless, the CF visual workload theory was effectively utilized to indicate the visual workload associated with aviator current proficiency.

The ICA subjects of the U-21 study were then compared to the IQA aviators from the helicopter instrument report. A representation of this comparison is referenced by Figure 6. Again, if the CF theory is an indication of cost or workload associated with the manipulation of a vehicle, the results would demonstrate that the UH-1H helicopter requires more visual workload on the primary Zone 1 instruments than did the U-21 fixed wing aircraft. These findings would be predicted by subjective data and the relative ratings of the stability of the visual data and the relative ratings of the stability of the two vehicles by other test agencies. However, the implications of the visual data are that if the helicopter stability was improved, then the crew could remain on station or in combat longer before becoming fatigued.

This same method of testing could be implemented to test future generation of helicopters to determine relative stability. If such aircraft did impose less visual input work to manipulate, they would provide a better platform for combat utilization.

To further expand the line of thought that the CF theory could reflect in some part the visual workload associated with the stability of the vehicle, a study has been completed. This investigation compared two groups of subjects with qualifications similar to the original SQA and IQA groups of the helicopter study (25). The two groups were; however, tested in an UH-1 flight simulator which was developed for the US Army to duplicate the flight, engine, and system characteristics of the UH-1 helicopter.

The results are summarized in Figure 7. The conclusions that can be drawn from these results are that the UH-1 simulator does, in general, have the same visual workload pattern as the UH-1 helicopter. However, because the visual workload in Zone 1 is higher for the simulator, the vehicle is less stable than the UH-1 helicopter. An expansion of the CF of Zone 1 can be seen in Figure 8. The three instruments that comprise this zone are indicated by AH for the artifical horizon, RMI for radio magnetic compass, and T-B for turn and bank indicator. From the major difference of the two vehicles as seen on the workload of the RMI, the indication would be that the UH-1 simulator is less stable mainly on the yaw axis. In addition, the inter-group subject differences in the simulator reflected the same results as had been reported in the UH-1 helicopter study.

CONCLUSIONS

In summary, this paper has attempted to address a method of assessing workload requirements imposed on one of five possible operator sensory channels. Since the demand of this visual input channel is estimated, as previously stated, to be 90% of the total input demands for vehicle manipulation, any theory which allows even a partial but precise description of the workload could aid future hardware designs, training, and mission delineation. These data will further be useful in determining an approach to reduce operator fatigue in the flight environment.

The current CF theory, although not the final answer, allows a more concise picture of visual workload than the classical methods which normally consist of the permutation of seemingly unrelated visual data points. The application section of this report demonstrated how the US Army Aeromedical Research Laboratory has collected and is continuing to expand a data base describing pilot visual performance in the military environment. Such data are considered invaluable to expand and test the current CF theory as well as providing an objective method to be utilized in answering current operational questions and problems. The examples were brief descriptions of studies which are already published in their entirety or are in the process of being completed. The implications from the results suggest that the CF theory is a valuable tool in testing and determining what the visual workload level should be for combat proficient pilots, how long pilots with varying degrees of proficiency could be expected to fly in the combat environment, and aircraft design requirements (such as stability), to reduce the onset of fatigue-induced errors. Additionally, the CF theory can be utilized to test and determine varying mission related workload, as well as the workload required by special equipment such as the night vision goggles, navigation equipment, and experimental flight displays.

The ability to measure visual input workload and/or psychomotor control is recognized as an invaluable tool required to validate instrument panel design, develop training and proficiency requirements and, in general, provide a more effective helicopter system for mission accomplishment.

TABLE 1

INSTRUMENT CLUSTERS WITHIN EACH ZONE

ZONE I	1.	Attitude Indicator	AH
	2.	Radio Magnetic Compass	RMI
	3.	Turn and Slip Indicator	T&B
ZONE II	1.	Altimeter	ALT
	2.	Airspeed Indicator	AS
	3.	Vertical Velocity Indicator	VSI
ZONE III	1.	Aircraft Monitoring Gauges	TORQ, RPM, ELEC, OIL, FUEL
	2.	Special Navigation Instrumentation	OBS
	3.	All Other Visual Areas	REST

REFERENCES

1. Senders, J. W., Fisher, D. F., and Monty, R. A. (Eds.) Eye movements and the higher psychological functions. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1978.

2. Hall, R. J., and Cusack, B. L. The measurement of eye behavior: Critical and selected reviews of voluntary eye movement and blinking (HEL TM 18-72). Aberdeen Proving Ground, Maryland: Human Engineering Laboratory, July 1972.

3. Klein, R. H. and Jex, H. R. An eye-point-of-regard system used in scanning and display research. Paper presented at the SPIE 15th Annual Technical Symposium, Anaheim, California, September 1970.

4. Monty, R. A. An advanced eye-movement measuring and recording system, American Psychologist, 1975, 30(3), 331-335.

5. Mowrer, O. H., Ruch, R. C., and Miller, N. E. The cornea-retinal potential difference as the basis of the galvanometric method of recording eye movements. American Journal of Psychology, 1936, 114, 423.

6. Milton, J. L., Jones, R. E., and Fitts, P. M. Eye fixations of aircraft pilots: Frequency, duration, and sequence of fixations when flying the USAF instrument low approach systems (ILAS) (USAF Tech. Rpt. No. 5839). Dayton, Ohio: Wright-Patterson Air Force Base, October 1949.

7. Fitts, Paul M., Jones, R. E., and Milton, J. L. Eye fixations of aircraft pilots: Frequency, duration, and sequence fixations when flying Air Force ground controlled approach system (GCA) (USAF Tech. Rpt. No. 5967). Dayton, Ohio: Wright-Patterson AFB, November 1949.

8. Mackworth, N. H. and Thomas, E. L. Head-mounted eye marker camera. *Optical Society of American Journal*, 1963, *52*, 713-716.

9. Frezell, T. L., Hofmann, M. A. and Oliver, R. E. *Aviation visual performance in the UH-1H, Study I* (USAARL Rpt. No. 74-7). Fort Rucker, Alabama: US Army Aeromedical Research Laboratory, October 1973.

10. Frezell, T. L, Hofmann, M. A., Snow, A. C., Jr. and McNutt, R. P. *Aviator visual performance in the UH-1H, Study II* Aeromedical Research Laboratory, March 1975.

11. Young, L. F. Measuring eye movements. *American Journal of Medical Electronics*, 1963, 300-307.

12. Sanders, M. G. Personal communication based on data being prepared for publication. November 1976.

13. Simmons, R. R., Kimball, K. A. and Diaz, J. J. *Measurement of aviator visual performance and workload during helicopter operations* (USAARL Rpt. No. 77-4). Fort Rucker, Alabama: US Army Aeromedical Research Laboratory, December 1976.

14. Siegel, A. and MacPherson, D. *Pilot opinion of the optimum arrangement of primary flight instruments in Naval aircraft* (NADA-AC-6910). Warminster, PA: Naval Air Development Center, September 1969.

15. Clark, W. and Itano, G. *Helicopter display improvment study* (IFC-TN-75-1). Randolph AFB, TX: USAF Instrument Flight Center, May 1975.

16. Simmons, R., Hofmann, M. and Lees, M. *Pilot opinion of flight displays and monitoring gauges in the UH-1 helicopter* (USAARL Rpt. No. 76-18). Fort Rucker, AL: US Army Aeromedical Research Laboratory, April 1976.

17. Diamond, S. Time, space and sterosccpic vision: Visual flight safety considerations at supersonic speeds. *Aerospace Medicine*, 1970, *41*, 300-305.

18. Sunkes, J. A., Pazera, E. E., and Howell, W. D. *A study of helicopter pilots' eye movements during visual flight conditions* (Task Assignment No. 59-205-10). Atlantic City, N. J.: Test and Experimental Division, September, 1960.

19. Stern, J. A., and Bynum, J. A. Analysis of visual search activity in skilled and novice helicopter pilots. *Aerospace Medicine*, 1970, *41*, 300-305.

20. Barnes, J. A. *Tactical utility helicopter information transfer study* (HEL TM 7-72). Aberdeen Proving Grounds, MD: Human Engineering Laboratory, April, 1972.

21. Barnes, J. A. *Analysis of pilots' eye movements during helicopter flight* (HEL TM 11-72). Aberdeen Proving Grounds, MD: Human Engineering Laboratory, April, 1972.

22. Simmons, R. R., Lees, M. A., and Kimball, K. A. *Visual performance/workload of helicopter pilots during instrument flight* (USAARL Rpt. No. 78-6). Fort Rucker, Alabama: US Army Aeromedical Research Laboratory, January, 1978.

23. Simmons, R. R. *Visual workload as a function proficiency* (Draft USAARL Rpt.). Fort Rucker, AL: US Army Aeromedical Research Laboratory. Report in preparation, October, 1978.

24. Simmons, R. R., and Kimball, D. A. *Aviator visual performance: A comparative study of fixed and rotary wing aircraft* (USAARL Rpt.) Fort Rucker, AL: US Army Aeromedical Research Laboratory. Report in preparation, October, 1978.

25. Simmons, R. R., Lees, M. A., and Kimball, K. A. *Aviator visual performance: A comparative study of a helicopter simulator and the UH-1 helicopter*. Paper presented at the NATO/AGARD Aerospace Specialists Meeting, Fort Rucker, AL, May 1978. (To be published in the Conference Proceedings.)

Figure 1. Visual Recording Equipment

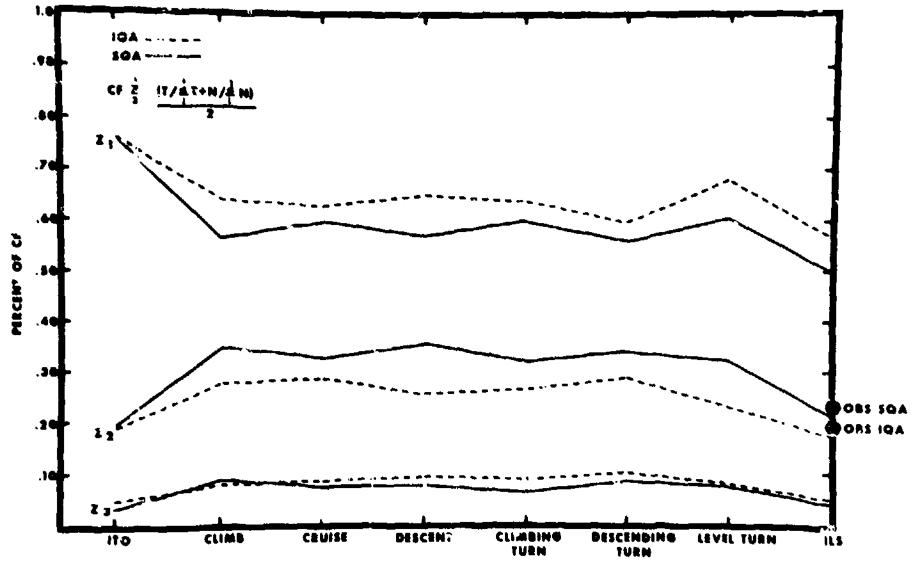


Figure 2. Graph of CF/Zone

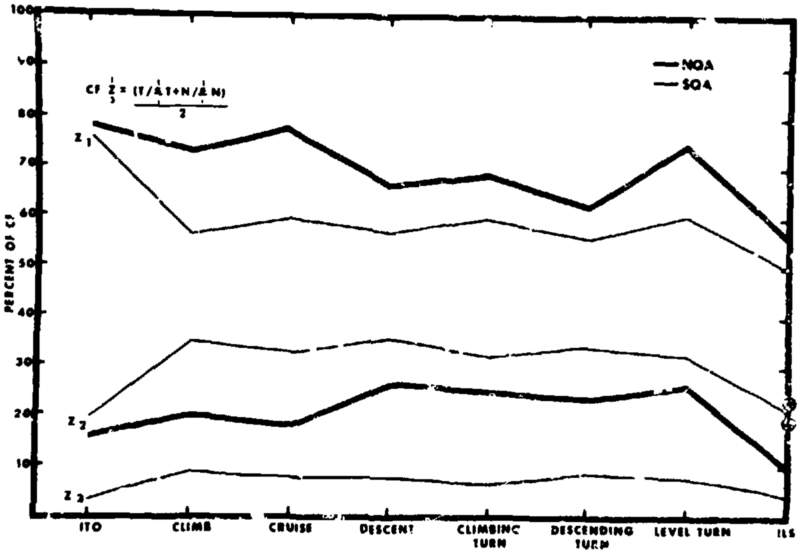


Figure 3. Graph of CF/Zone

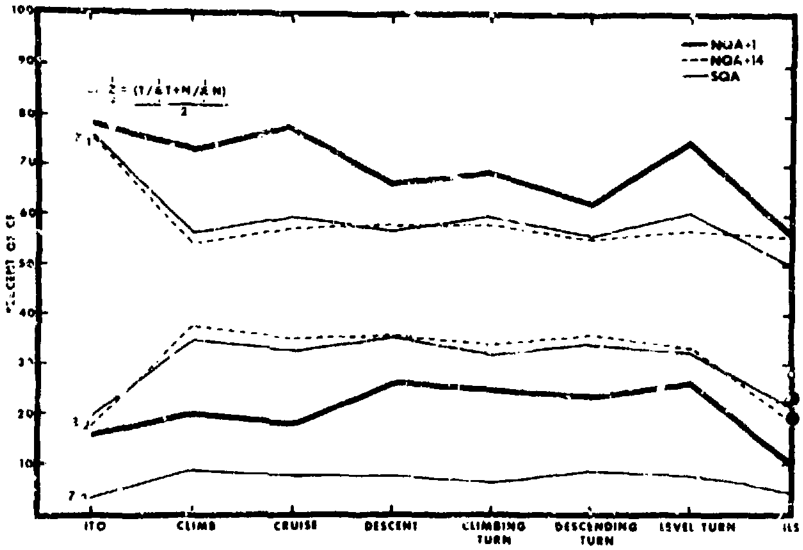


Figure 4. Graph of CF/Zone

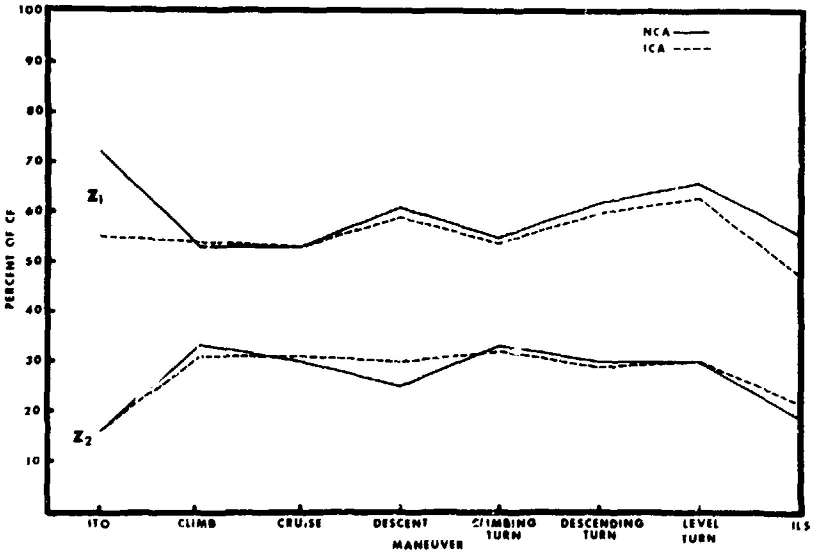


Figure 5. Graph of CF/Zone—Fixed Wing Aircraft

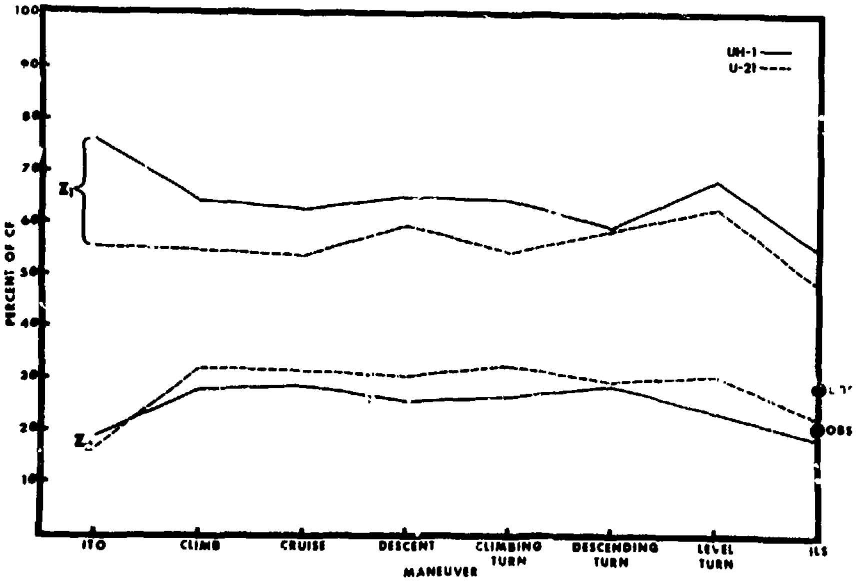


Figure 6. Zone/CF for IZQ/Aircraft

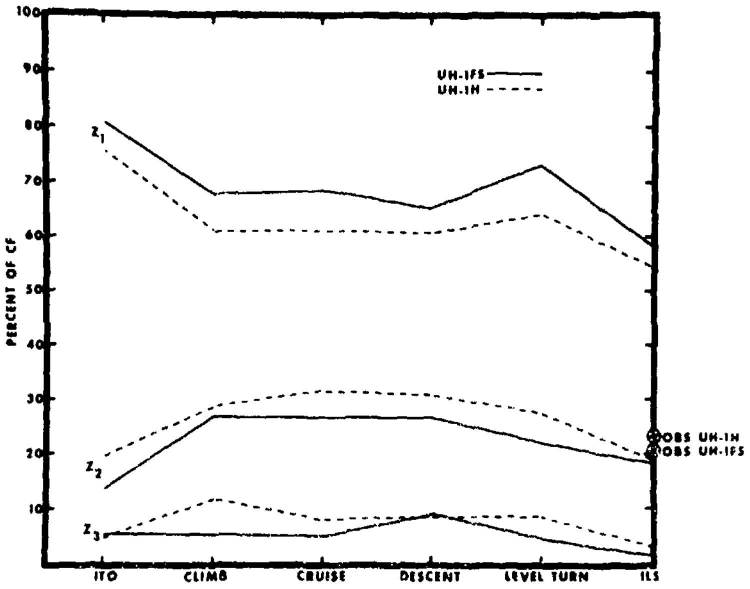


Figure 7. Graph of CF/Zone

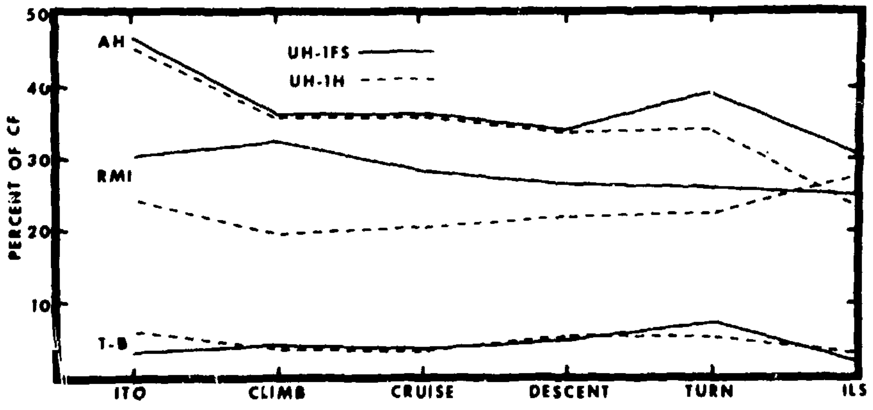


Figure 8. Graph of CF/Zone 1

HANDLING QUALITIES, WORKLOAD AND HEART RATE

by

Alan H. Roscoe
Royal Aircraft Establishment
Bedford England

Introduction

The important and close relationship between aircraft handling qualities and pilot workload has been underlined by several authors. Twelve years ago Westbrook and his colleagues (1) stated: "To a pilot the multiple stresses of flight, his workload, are summerized under his judgement of the handling qualities." Today, there is an increasing tendency for the pilot to be less of an active controller and more of a supervisor, but even so, this statement - especially when applied to short term workload - still holds true. This changing role of the pilot has led to a wider interpretation of the term handling qualities. Cooper (2) remarked: "Because of preoccupation with manual control in the past, it has been a general practice to associate handling qualities primarily with aircraft stability and control characteristics. Actually, handling qualities encompasses not only the aircraft stability and control but the total of the pilot-aircraft interface features as well ....." Most people now interpret the term handling qualities in this way and it is convenient to do so in this paper.

Unfortunately, no such agreement exists about the interpretation of the term workload. It is, therefore, important for authors to make clear their own interpretation of the term. In this paper, the basic idea of pilot workload is considered to be effort-related, as distinct from task-or performance-related concepts. A suitable definition is that given by Cooper and Harter (3): "the integrated physical and mental effort required to perform a specified piloting task." The idea of workload as effort is one with which most pilots would agree (4); and it is consistent with the measurement of heart rate as a means of assessing workload.

Assessing handling qualities and the associated workload is an important part of flight evaluation, whether of control and stability or of guidance systems, and various assessment methods are used by test pilots and engineers for this purpose. Measurement of performance, which is an important and essential part of control and guidance evaluation, may be used to estimate changes in both handling qualities and workload (5) (6). Unfortunately, changes in handling and workload are not always refected by changes in performance. In 1956, Duddy (7) highlighted the difficulty of estimating the extent of improved stability in a directionally unstable fighter when fitted with a yaw damper as aiming accuracy was not improved. As Spyker et. al. (8) have observed: "An evaluation procedure which relies exclusively on performance measures is inadequate. That is, a pilot with one configuration may work twice as hard as he does with another, yet achieve equal performance with both." This ability of pilots to "compensate" is referred to by Cooper and Harper (3) in their Handling Qualities Rating Scale.

Another method of assessing handling qualities and levels of workload, especially during landing approaches, is by measuring control activity. Morrison and Stimely (9) quantified pitch activity and used the results to augment pilot's subjective impressions of workload during noise abatement approaches. Barber et. al. (10) summated force inputs from elevator, aileron, and rudder to give a workload factor during the evaluation of general aviation aircraft handling qualities. Nevertheless, these authors accepted that using force inputs to give a workload factor ". . . has some deficiences."

Objective techniques, especially if they involve precise measurement, are particularly attractive to engineers. However, by far the most used techniques for evaluating handling qualities and workload are subjective. These techniques, which vary from simple comments by pilots to complicated questionnaires and rating scales have, for the most part, been developed for assessing aircraft handling rather than pilot workload. A well known and accepted handling qualities rating scale is that of Cooper and Harper (3), which refers to workload by asking the question: "Is adequate performance attainable from tolerable workload?"

Clearly, workload levels for a given task are related to the aircraft's handling characteristics, but a valid rating for the latter may not always give a reliable estimate of workload. Experienced test pilots may be quite adept at using opinion rating scales but occasionally it seems difficult to separate assessments of workload from those of handling qualities, leading to anomalies and ambiguities. Westbrook and his colleagues (1) commented that: "If a reliable method were available to obtain a measure of workload or stress, it is undeniably true that many of the anomalies in handling qualities data could be explained."

Several investigators have recorded physiological variable from pilots in real and simulated flight as a means of estimating levels of stress and workload. This paper, by describing two current flight trials and by referring briefly to previous studies, examines the relationship between pilot's heart rate and subjective assessments of handling qualities and workload.

Materials and Methods

All the subjects referred to in the following examples were qualified test pilots who were experienced and current on aircraft type. Most of the flight trials involved either the take-off or the approach and landing and so the task was well defined and realistically demanding. Performance was closely monitored by on-board instrumentation and by airfield sited kinetheodolites. Whenever possible flight trials were designed in such a way that experimental variables could be compared during the same sortie. In this way the effects of weather, learning and other irrelevant influences were minimised.

Various aircraft, ranging from pure research to representative civil and military types, were used

Pilot assessments of handling qualities and workload were made by using the Cooper-Harper scale, by straight-forward comments, or by a questionnaire designed for a particular trial. In most cases the pilot recorded his comments or gave a rating while in the aircraft; questionnaires were completed after landing. Latterly, a formal workload rating scale, based on the Cooper-Harper handling scale, has been constructed and is currently being evaluated.

At Bedford, pilot's heart rates are obtained by recording the ECG signal in analogue form with the "R" wave being used to trigger a cardiotachometer. The resulting beat-to-beat rate is then plotted against time for initial examination and analysis (Fig. 1). Subsequently, mean heart rates for consecutive 30 sec epochs and mean values for a particular flight phase or sub-phase are used to compare levels of workload.

Flight Trials

1. "Ski-jump" Ramp take-offs.

The first example is of a trial to assess the advantages of using an inclined ramp to improve the take-off performance of shipborne H S Harrier VTOL combat aircraft (11). The aircraft is accelerated on to the ski-jump shaped ramp from a short run (50-100m) with nozzles rotated rearwards. At the top of the ramp, and on the point of becoming airborne, the nozzles are rotated downwards to a pre-set angle. After reaching conventional flying speed the nozzles are rotated back to the aft position.

One of the Harriers used in the trial, a two-seat version (T3), is equipped for telemetering heart rate from both cockpits. A portable EEG recorder (Oxford Instruments) is used to monitor pilot's heart rate from the other (single seat) aircraft.

Two or three ramp take-offs are normally carried out on each sortie with the take-off weight and distance being varied between runs. The ramp was set at an angle of 6° for the first series of take-offs and at 9° for the second series. Incremental increases in ramp angle are planned for future stages of the trial.

Performance measurements are essential in this trial but assessments of handling and workload, especially for the critical period of accelerating along the ramp during the partially jet-borne flight phase, are also important. Pilot's heart rates together with subjective ratings of handling qualities, using the Cooper-Harper scale, and of workload, using a special ten-point rating scale have been obtained for a large number of take-offs.

Three test pilots, flying the T4, proved the bulk of the heart rate data and subjective ratings. Take-offs with and without autostabilisation, during simulated and real night conditions, and in crosswinds up to 10K were evaluated.

Table 1 gives overall pilot ratings, and mean heart rates for 40 sec epochs centred on the time of nozzle rotation. This epoch includes a period of from 10 to 15 sec before releasing the brakes, a time when pilots carry out a final check of instruments and configuration.

TABLE 1

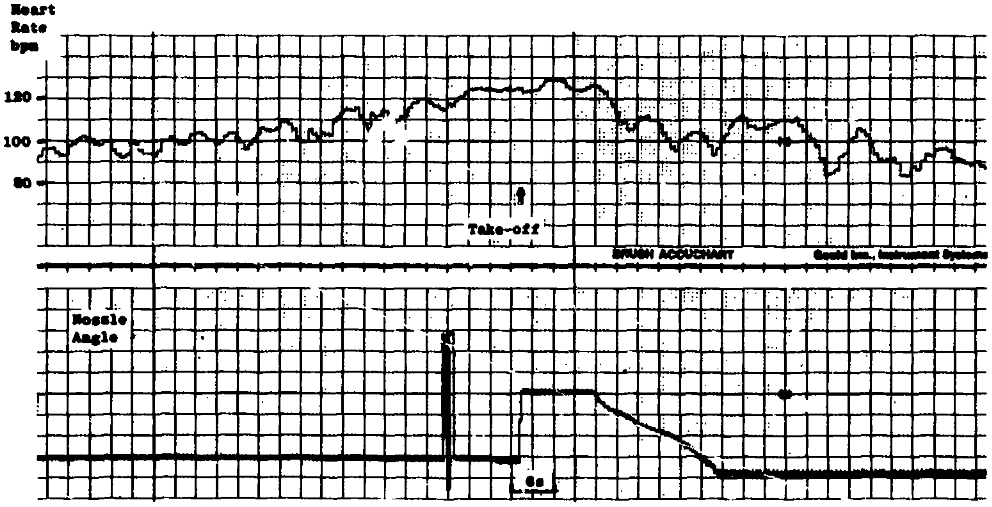
6° HARRIER RAMP TAKE-OFFS ("NORMAL" CONDITIONS). MEAN HEART RATES (40s), HANDLING RATINGS (COOPER-HARPER) AND WORKLOAD RATINGS.

Pilot	n =	Heart Rate bpm	HQ Rating	WL Rating
A	15	110.9	4	4
B	11	119.3	4	4
C	5	93.6	3	4

Fig. 1 is a typical beat-to-beat heart rate plot for the handling pilot during a ramp take-off flown from the front seat. Close examination shows that his heart rate increased some 10-20 sec before going to take-off power prior to releasing the brakes. Pilot comments confirm that the workload increases rapidly at this time and remainds high until conventional flight some 30 to 40 sec later. Overall assessment of handling qualities and workload, after the 6° stage of the trial, were favorable and ramp launches were considered to be easier than normal runway short take-offs (STOs).

Comparison of mean 60s heart rate shows no difference between types of take-off if the epoch includes the 15-20s before rolling. However, if the epoch starts when brakes are released the runway STO heart rates tend to be 2-3 bpm higher (Table II). The finding agrees with ratings for handling qualities and for workload. The influence of ground effect, during runway STOs, seems to cause a deterioration in handling with a consequent increase in workload.

Neither night take-offs, both simulated and real, nor crosswinds up to 10K caused any difficulty and resulting heart rate values and ratings were similar to those for 'normal' ramp launches. Take-offs in the unstabilized mode tended to result in higher ratings with marginally increased heart rates. Results of the 9° ramp evaluation showed ratings of handling qualities and workload to be similar to those for 6°. Heart rates, which were slightly lower, agreed with pilot opinion that 9° ramp take-offs were no more difficult and could well be easier than those at 6°. Pilots commented on a smoother ride along the 9° ramp.



HARRIER "Ski-Jump" Take-off

FIGURE 1. HS HARRIER. BEAT-TO-BEAT HEART RATE AND NOZZLE ANGLE RAMP TAKE-OFF

TABLE II

COMPARISON OF HARRIER STO MEAN HEART RATES (60s)

	6^{o} Ramp		Runway	
Pilot	1	2	1	2
A	110.8 (n = 15)	108.9	109.3 (n = 21)	110.8
B	116.5 (n = 11)	114.8	115.9 (n = 6)	117.3
C	90.4 (n = 5)	89.6	90.8 (n = 18)	91.2

1 Epoch from 15 - 20s before releasing brakes.

2 Epoch from releasing brakes.

2. Direct Lift Control.

A modified BAC 1-11 is currently being used to evaluate the benefits of using Direct Lift Control (DLC) to improve handling and performance during the approach and landing. DLC should enable the aircraft to be flown more precisely on the glide slope and also result in better all round landing performance with less touchdown scatter. Workload during this phase should be reduced and the ability to cope with turbulence improved. More direct control of descent rate should prove beneficial during steep gradient approaches and improve safety during the flare.

In addition to monitoring aircraft performance, pilot assessment of handling and workload, and measurement of pilot's heart rates, are important trial requirements.

The DLC system fitted to the 1-11 uses the four wing spoilers to generate lift changes; these are controlled by electrical sensing of control column pitch movements. A wash-out circuit is inserted into the system to provide the pilot with relatively normal acceleration responses to pitch inputs.

The first batch of flying (phases one and two) was concerned mainly with optimising the control characteristics and giving the pilots some experience of new handling techniques. Control was rated better after the DLC was [illegible] linear [illegible] better still after a lag was incorporated in the DLC signal from the control column. Pilot [illegible] monitored on only a few sorties during this stage of the trial.

The main flight investigation (phase three) was aimed at evaluating DLC during both conventional 3° approaches and 6° steep gradient approaches. Each sortie included a batch of basic-aircraft runs for comparison with DLC.

For most sorties pilots were briefed to concentrate on a precise position at 50 ft and glide slope tracking was not appreciably better. But on the occasions when pilots were briefed to maintain precise glide slope tracking, performance improved. This was particularly evident on the 6° approach gradient. Landing performance was definitely better when off steep approaches but not noticeably different when from 3° approaches. The ability of DLC to quickly arrest descent helped to produce more accurate and smoother landings, but if the flare was started too early there was a tendency for pilot induced oscillations (PIOs) to occur.

Overall pilot assessments indicated that the aircraft's handling qualities in the landing configuration were improved with DLC. Workload during the approach and in the flare was thought to be generally lower, but especially so for the 6° glide slope. Heart rate responses appeared to agree with pilot opinion although, at first there were a small number of discrepancies. These were resolved after discussion with the pilots. For example, when PIOs occurred heart rate responses for the flare epoch were much higher; the overall effect was to result in mean values for this manoeuvre which were similar whether DLC was used or not.

Perhaps it is not surprising that pilot ratings for the flare, both of handling and of workload, varied considerably according to whether PIOs were present or not!

Fig. 2 shows mean heart rate values for 3° and 6° approaches, with and without DLC, flown in similar weather conditions by one of the three project test pilots. These results show an obvious trend in favor of DLC but the only appreciable reduction in heart rate is during the glide slope interception and early part of the 6° approach.

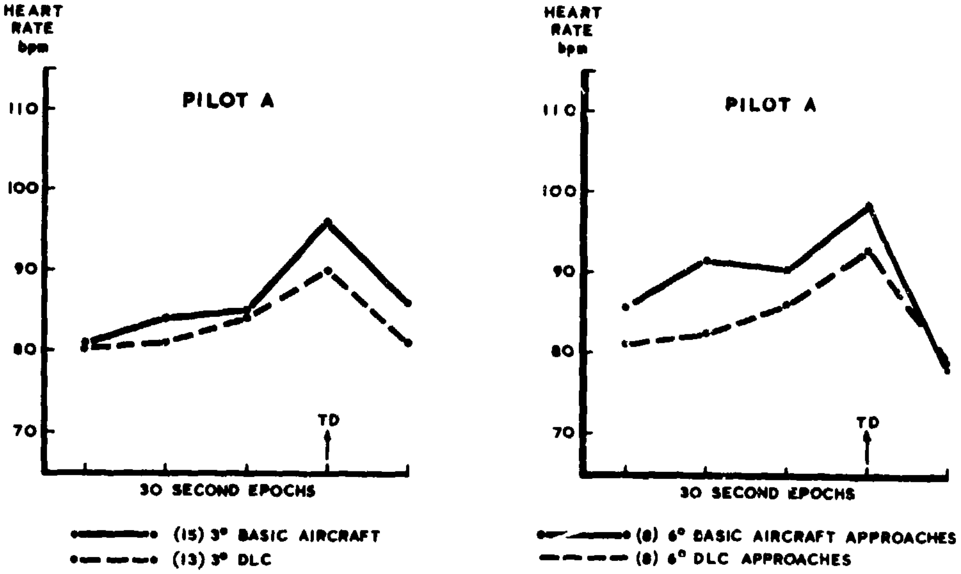


FIGURE 2. BAC 1-11. MEAN 30s HEART RATE VALUES FOR DLC AND BASIC-AIRCRAFT

Results for one of the other two pilots were similar to those illustrated, again showing a definite trend in favor of DLC. Mean heart rate values for the third pilot did not differ appreciably between DLC and basic aircraft approaches. In fact, because PIOs seemed to disturb this pilot more mean rate for the flare from 6° approaches was slightly higher with DLC. He assessed handling qualities overall as being improved with DLC, he was unsure about workload on 3° slopes but felt it was reduced on 6° approaches.

Results from subsequent flight (phase four), following minor changes to the system, have confirmed the benefits of DLC. However, because pilots were briefed to fly more accurate glide slopes, heart rate values were not noticeably reduced, but performance was improved. A sortie of 3° and 6° approaches flown in turbulence provided the opportunity to demonstrate the advantages of DLC. This was confirmed by the markedly lower heart rates for DLC approaches when compared with basic - aircraft approaches.

It was hoped to carry out sufficient flying to allow statistical analysis of results, but the number of sorties has been limited and only trends were established.

Comments

The two trials described above are both typical examples where assessment of handling qualities and related workload are important features. However, they differ in some respects. The concept of the 'ski-jump' ramp is aimed at increasing the maximum take-off weight of ship-borne vectored thrust combat aircraft thereby improving their overall tactical performance. This being the primary objective. Handling qualities and workload are of secondary importance; it is only necessary to ensure they are not increased beyond a level which might jeopardise the take-off.

DLC, on the other hand, is aimed primarily at improving safety during the approach and landing. Therefore, assessment of handling and workload assumes much greater importance. Of course, performance, because of its relationship to safety, is also important.

In both examples there is generally good agreement between handling qualities, workload, and heart rate. The few anomalies that have occurred, especially in the early stages of the trials, have been resolved by detailed discussion with the pilots. For example, a high heart rate and high workload rating but low rating for handling qualities during the first ramp take-off by one test pilot was due to excessive anticipation. The pilot rated the workload as 7 (out of 10) and the handling as Cooper-Harper 3. His 40 second mean heart rate was 156 bpm. Afterwards he reported: "with the benefit of hindsight, I realize that I was much more keyed up than I need have been, and I expect that my workload will be very much less on subsequent launches as I gain experience: His subsequent workload ratings averaged 4 with autostabs on and 5 with autostabs off and the corresponding handling assessments were, similarly, Cooper-Harper 4 and 5. The overall mean heart rate level for eleven 6° take-offs was 119 bpm. The high heart rate generated by the first ramp take-off is typical of the increased arousal experienced by test pilots about to carry out a novel flight task. Roscoe (12) has suggested that experimental test pilots frequently overestimate the level of difficulty for the first run of an untried or unusual manoeuvre or task.

Occasional heart rate measurements of two pilots during the preliminary phases of the DLC trial resulted in appreciably lower values for the new system compared with the basic-aircraft. These pilots were also most enthusiastic in their early comments on the system. It was, therefore, something of a disappointment to find smaller decreases in heart rate when it was routinely monitored during phase three. Subsequent discussion revealed that in flight trial proper, pilots had changed their strategy and flew more precisely than in the previous stage. Improvements in handling were apparently being used to increase performance although this was not always measurable. Occasional discrepancies between heart rate and pilot opinion were caused by failure to record the fact that PIOs occurred during the flare.

Previous Studies

For some nine years, at RAE Bedford, pilot's heart rates have been monitored during various flight trials as part of a long term study of workload and stress. Evaluation of handling qualities was a primary requirement of many of these trials and it is interesting, and perhaps profitable, to refer briefly to some earlier ones.

Autostablisation systems should lessen the effects of poor stability and control and thus lead to an overall improvement in handling. The VTOL research Short SC1 was an example of a relatively unstable aircraft and pilots comparing the stabilised with the unstabilised configuration invariably commented on the marked improvement in handling qualities of the former. It is interesting to compare heart rate responses for one pilot flying two similar 6 min sorties consisting of a vertical take-off to 30m (100 ft), small accelerations and decelerations, ending in a vertical landing. The first flight was stabilised and resulted in a mean heart rate of 109.6 bpm.: the second flight, which was unstabilised, resulted in autostabilisation, but similar heart rate comparisons for other test pilots did not show any differences. Detailed discussion with pilots revealed that most of them suspected the integrity of the autostabilisation system and transferred the spare effort made available by improved handling to monitoring the system itself.

Even though handling qualities and workload are closely related, it does not necessarily follow that an improvement in handling will invariably lead to a reduction in workload. It may sometimes be perferable, especially for well motivated pilots, to improve performance and maintain the same level of workload. If performance is monitored such improvement will be obvious and this alone will indicate the defree of benefit gained by improved handling. Such improvement is evident in some of the DLC flying referred to earlier. Pilots also make use of additional spare effort or capacity to increase monitoring or to carry out other covert tasks which are not immediately obvious.

Gerathewohl (13) made the point that subjective ratings of handling qualities ".... as accurate as they may be in regard to control desirability or difficulty, do not contribute to workload determination, since they are only loosely connected to task demands and pilot response." Certainly, as in the above example, a pure handling qualities scale may not give an accurate estimate of workload. It is clear that subjective assessments of workload must be derived from rating scales specifically designed for the purpose.

Subjective assessments, in general, as sometimes unreliable, for example, it is known that they are susceptible to both inter- and intra-subject inconsistency. In particular, subjective ratings for what may be minimal changes in handling characteristics can be misleading by suggesting the existence of larger differences, especially if the ratings were obtained under different conditions. Such anomalies may be due to a poorly designed rating scale, because the test pilot has varied his assessment strategy, or because of undetected changes in flight conditions.

This problem is typified in the trial of a powerful rudder autostabiliser fitted to the BAC 221 slender delta supersonic research aircraft. Lateral directional handling characteristics during the landing approach were assessed by three test pilots. The task consisted of a "side-step" manoeuvre at a height of 75m (250ft) placing the aircraft to one side of the centre line. An "S" turn was necessary to realign the aircraft with the runway, thereby testing the effectiveness of the system. Different autostabiliser settings were evaluated and, as it was possible to vary these in flight, the associated handling characteristics were compared under similar conditions. The Cooper-Harper rating scale was used for this purpose. The pilot's heart rate was monitored on several flights so that mean values for each approach could be compared. It was also possible to examine the relationship between levels of heart rate and ratings of handling qualities.

Except for the extreme autostabiliser settings and for "no autostab" approaches when ratings and heart rates were appreciably higher, results were disappointing. Heart rate values were inconsistent, tempting, therefore, to conclude that the differences between the various autostabiliser settings were inconsequential and that heart rate measurement correctly interpreted this fact.

Disagreements between workload assessment and heart rate, which have been rare, have tended to occur during relatively undemanding tasks when changes have been minimal or unimportant.

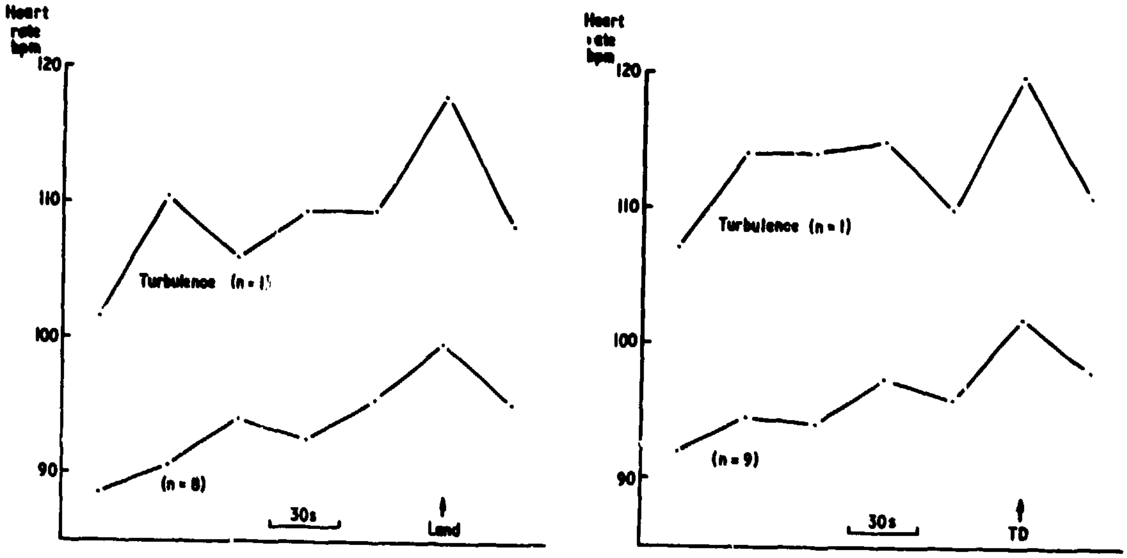


FIGURE 3. VC-10. MEAN 30s HEART RATE VALUES FOR NORMAL CONDITIONS AND SEVERE TURBULENCE. a. 3°, b. 5°/3°, APPROACHES AND LANDINGS

Different weather conditions can influence handling to varying extents, and turbulence, in particular, causes increased workload by degrading stability and control. This is especially noticeable during a flight task where accurate tracking is required. Fig. 3 compares the heart rate responses of a pilot flying two different types of approaches and landings in severe turbulence with mean rates for similar approaches flown in relatively smooth conditions. This example is from a flight trial of noise abatement approaches using a BAC VC - 10, (14). It can be seen that there are marked increases in heart rate for both types of approach though the increase is marginally greater for the earlier and steeper section of the 5°/3° two-segment profile when compared with the 3° gradient. These findings agreed closely with the pilot's assessment of the changed handling qualities and workload. He considered that turbulence increased the workload more for the two-segment than for the conventional approach, especially during the acquisition and early part of the glide slope.

Pilots occasionally reveal some degree of bias towards or against a particular experimental flight condition which, based on falacious reasoning, may affect their judgement and result in misleading subjective ratings.

In the early stages of a series of flight trials to evaluate various types of noise abatement approaches initial pilot opinions of 7½°/3° two-stage flare approaches, in a HS Andover, were unfavorable Pilots felt instinctively that transiting from a 7½° slope to one of 3° at a height of 200 ft would be too demanding and they rated the workload quite high. However, from the beginning of the trial heart rate responses for this approach profile were similar to those for convential 3° approaches. Careful thought and subjective re-analysis, by the two test pilots who flew early sorties, led to a review of their original subjective assessment of workload. Subsequently, these pilots, together with other participating pilots, tended to prefer the 7½°/3° approaches to the 3°. They considered that improved handling on the steep segment was an important factor in maintaining a reasonable level of workload. Heart rate and workload assessments showed good agreement (15).

In this instance both the task and the handling qualities are changed but the net result is that workload and heart rate are unchanged.

The previous trials were concerned mainly with stability and control which to many people used to be the accepted interpretation of the term handling qualities. But, as stated earlier, evaluation of guidance systems is also relevant; indeed a large proportion of test flying at Bedford is directed to evaluating approach guidance displays and systems.

A typical trial, which took place in 1971, was to evaluate an airborne visual approach indicator (VASIO presented as a HUD. For the purpose of the experiment only omnidirectional runway edge lights and green threshold lights were used. All other lights were estinguished and moonless nights were selected for the trial sorties. Two pilots alternated as experimental pilot (P1) and co-pilot (P2) for four sorties giving a total of 32 approaches. Each run ended in an overshoot at 30m (100 ft). Four difference runways were used in order to reduce any element of learning, though, as it happened, varying weather conditions eliminated this effect. The aircraft used for this stage of the trial was an HS Comet 2E.

Glide slope performance was considerably better with the HUD-VASI and the heart rate of the handling pilot was reduced. Compared with no-aid approaches, the overall decrease was 4.2 bpm for one and 6.8 bpm for the other pilot; whereas there were only negligible differences in heart rate when acting as co-pilot. Both pilots were keen to point out that workload was significantly reduced by the HUD and they anticipated larger decreases in their heart rates but agreed that the benefit of the aid would have been greater had the experimental approaches ended in a landing rather than in an overshoot.

Improvement in performance without any evidence of increased workload was, in itself, adequate proof of the advantages of the HUD-VASI. Nevertheless, the trial scientists were delighted to have evidence of a reduced workload as well. Unfortunately, because of the severe and widely differing weather, the wide variations in heart rate between the sorties precluded statistical significances.

These studies, which have used examples of data obtained during operational test flying, demonstrate the use of heart rate as an indicator of workload. Such data have proved to be of great value in the overall study of pilot workload, but it has to be admitted that the direct value of heart rate measurement in evaluating handling qualities is still not clear. Nevertheless, pilots and engineers associated with these trials consider monitoring heart rate to be a worthwhile adjunct to those techniques commonly used in flight evaluation. It should be noted that these examples have been confined to trials where the pilot has been handling the controls. Heart rate changes for handling pilots have, for the most part, proved to be reliable indicators of important changes in workload when the task has been realistically demanding. But in other trials at Bedford, where the pilot has been in a monitoring role his heart rate responses did not appear to reflect changes in workload with anything like the same reliability. This difference in heart rate sensitivity, between the pilot in the control loop and the pilot outside the loop, is important.

Discussion

A large number of reports on aircraft handling trials refer to related levels of pilot workload. However, it is patently obvious that in most instances assessing handling characteristics was the primary objective whereas estimating workload was very much a secondary aim. This approach is usually adequate and leads to realistic estimations of workload, but it is apparent that sometimes a pilot's main concern with handling has adversely affected his ability to assess workload. Ellis (16), in pointing out that it is important that ratings for workload and handling qualities are not confused, wrote: "When pilots are asked to make a formal assessment of workload as a primary measure, it should be absolutely certain that workload is the ultimate aim of the exercise." Ellis also observed: "Workload is always important in handling qualities investigations and so pilots should be encouraged to comment on it and rate it but workload should not be allowed to usurp the place of the handling qualities rating where the latter is the more appropriate measure."

It is clearly an advantage to use a specially constructed rating scale for assessing workload during flight testing. Unfortunately, such scales suffer from the same problems and attract the same criticisms as do pilot opinion scales for assessing handling qualities. By using some other method of estimating workload it may be possible to augment pilot opinion and, perhaps occasionally, resolve anomalous findings. Physiological variables, which have been recorded by many research workers during studies of pilot workload and stress, may be used for this purpose. The literature, though, contains few reports where the relationship between handling qualities, workload, and physiological responses has been studied in detail.

In 1962 Roman and Lamb (17), in discussing the results of measuring heart rate in flight, observed that: "Pulse rates correlate well with the pilot's estimates of the difficulties connecting with handling the aircraft during any one phase of flight." Rowen (18) pointed out that high heart rates recorded from the pilot of the M2 lifting body were associated with the poor lift/drag characteristics, which made particularly heavy demands on pilot skill. By measuring pilot's heart rates, Billings <u>et. al.</u> (19) demonstrated that helicopters fitted with hydraulic boost systems were significantly less demanding to fly. Hasbrook and his co-workers (20) used heart rate measurement to augment pilot opinion during the flight evaluation of a new instrument display. They were able to show that the new display, which reduced panel space by 25%, was an acceptable alternative to the conventional display.

These examples from the literature demonstrate a relationship between handling characteristics and workload as indicated by heart rate. But what is the extent of this relationship? Is it reliable and consistent? Can it be usefully employed in flight evaluation? Is monitoring pilot's heart rate during test flying a practicable exercise?

The technique of monitoring heart rate is relatively simple, it does not intrude into the flight task nor does it compromise flight safety. It is readily accepted by pilots. In fact, Bedford test pilots have co-operated to the extent of applying their own electrodes and preparing their monitoring equipment for flight on many occasions. The resulting heart rate data are often studied with interest by the pilots who find them helpful in recalling various aspects of the sortie.

Heart rate does not give absolute values of workload and so in order to obtain meaningful results it is necessary to use it as a comparative measure. It is worth noting that pilot rating scales, though appearing to give absolute values, because they are subjective, are really scales of comparison as well (21).

The examples of flight trials presented in the previous section relied on comparison with another experimental condition or with some form of datum; and wherever possible the comparison was made during the same sorties. The trials by Billings and his colleagues (19), and by Hasbrook <u>et</u>. <u>al</u>. (20), referred to above, were similarly based on comparison.

To compare the changes in workload caused by different handling characteristics it is necessary to ensure that other aspects, such as the flight task itself, remain constant. This is sometimes difficult to achieve. For example, different flap settings on the approach may change handling but may also require

Subjective assessments of handling and workload appear to be more consistent when they demand a high level of piloting skill than when they require little effort. Likewise, heart rate measurements are more consistent and reliable if the experimental flight task if realistically demanding. Data from the Bedford studies show that test pilots have, for the most part, given estimates of workload which agreed well with their heart rate levels. But the agreement was better when heart rates, and presumably workload levels, were higher; and as might be expected, anomalies tended to occur more often when workload and heart rates were low. It is interesting to note, though, that from measurement of heart rate and finger tremor, Nicholson and his co-workers (22) concluded that: "..... high workload associated with difficult approaches and landings rendered the pilot's subjective assessment more variable."

Results from workload studies made in real flight are generally more reliable than those made in simulated flight and this must be particularly so in studies of handling - related workload. Unfortunately, it is difficult to set up well controlled flight experiments and so there is a strong temptation to resort to using laboratories and simulators for workload investigations. Protagonists of these techniques point to the undoubted value of research simulators in assessing handling qualities. But laboratory and simulator experiements tend to restrict the number of input parameters to which the pilot is assumed to respond. In real life the pilot is faced with a wide range of input information - much of it redundant, but all liable to have some effect on his behavior and hence his workload.

As noted in the HUD trial, the inability to control such variables as weather, and the small number of experimental sorties - limited by the high cost of flying aeroplanes - often results in differences in heart rate which are not statistically significant. Nonetheless, trends, especially if they support pilot opinion, may be quite adequate, but even if they conflict, heart rate data can be most valuable in attracting attention to possible ambiguities. Further examination and discussion with the pilot may then reveal previously undetected factors. Beat-to-beat heart rate is particularly useful in identifying short term changes in workload which may not be obvious to a pilot making an overall assessment.

Unfortunately, most of the flight trials at Bedford did not use numerical rating scales for assessing workload and some trials did not use them for assessing handling. These omissions, together with the limited number of experimental sorties, has precluded any opportunity for statistical analysis of the relationship between handling qualities, workload and heart rate. A flight trial designed to examine this relationship more closely is currently underway. Three test pilots, flying three different aircraft, compare handling characteristics during various demanding tasks. These include the approach and landing, low-level high speed flight, and formation flying. Pilots use the Cooper-Harper scale for rating handling qualities and a 10 point scale (based on Cooper-Harper) for rating workload. Heart rate is recorded on all sorties.

It is hoped that this investigation will result in enough data to permit some degree of statistical analysis. It is worth noting, though, a point made by McGregor (23) who stated: "One of the criticisms of numerical pilot rating scales as opposed to adjectival scales is that statistical games will be played with numbers that are not statistically meaningful." He continued: "If statistical indices are used they must be adequately enough defined to enable the reader to assess their validity and sufficient data presented to allow a check to be made of the results." With this in mind, it is not intended to attempt to identify any mathematical relationship between assessments of handling or workload, based on rating scales, and heart rate. The individuality of pilots makes this virtually an impossible task anyway.

Summary and Conclusions

There is obviously a distinct advantage in augmenting a pilot's subjective assessments of handling qualities and workload. This paper presents practical examples where pilot's heart rates have been used to augment their opinions of handling and workload during various flight trials. These studies provide good evidence to show that, in general, this technique gives reasonably good indications of the workload generated by particular handling qualities.

The technique of monitoring heart rate is simple, it is accepted by pilots, and it is compatible with test flying. To improve reliability and consistence the flight task should be realistically demanding and require the pilot to be in the handling loop. Comparisons between experimental conditions, or with some form of datum, give more meaningful results; wherever possible comparisons should be made during the same sorties. Raw data in the form of beat-to-beat heart rate are invaluable for revealing rapid and short duration changes in handling qualities which affect workload.

In this way, potentially misleading results can be identified in good time, thereby drawing attention to the need for further investigation. Anomalous findings may be resolved by examination of heart rate data and by discussion with the pilot.

The author has not made any attempt to satisfy strictly scientific criteria, the primary objective being to draw attention to the value of using heart rate as a flight test procedure. But in addition, it is hoped to stimulate thought and discussion so that it may be possible to reduce some of the anomalies found in handling qualities and workload, referred to by Westbrook et. al. (1).

REFERENCES

1. Westbrook, C. B., Anderson, R. O. and Pietrzak, P. E. Handling qualities and pilot workload, Conference Proceedings No 14 Assessment of skill and performance in flying, AGARD, Paris, 1966.

2. Cooper, G. E. The pilot-aircraft interface, Symp Vehicle Technology for civil aviation, seventies and beyond, NASA SP - 292, Washington DC, 1971.

3. Cooper, G. and Harper, R. P. The use of pilot rating in the evaluation of aircraft handling qualities, NASA Tech Note TNXD - 5753, Washington DC, 1969.

4. Ellic, G. A. and Roscoe, A. H. Pilot workload: A survey (In preparation).

5. Brickson, C. A. Pilot landing performance under high workload conditions, Conference Proceedings No 146, AGARD, Parris, 1974.

6. Lees, M. A., Kimball, E. A. and Stone, L. M. The assessment of rotary wing aviator precision performance during extended helicopter flights, Converence Proceedings No 217 Studies on pilot workload, AGARD, Parris, 1978.

7. Duddy, R. R. The quantitative evaluation of an aircraft control system, AGARD Report No 29, NATO, Paris, 1956.

8. Spyker, D. A. et. al. Development of techniques for measuring pilot workload, NASA CR-188, Washington DC, 1971.

9. Morrison, J. A. and Stimely, R. L. An operational look at the two-segment approach, AIAA Paper No 74-979, 1974.

10. Barber, M. R. et. al. An evaluation of the handling qualities of seven general-aviation aircraft Technical Note, NASA TN D-3726, Washington DC, 1966.

11. Coleman, H. J. Harrier 'Ski Jump' launch studied, Aviation Week and Space Technology 105, pp. 17-18, December 1976.

12. Roscoe, A. H. Stress and workload in pilots, Aviation Space Environ Med (In press).

13. Gerathewohl, S. J. Definition and measurement of perciptual and mental workload in aircrew and operators of Air Force weapon systems, A status report, AGARD CP-181: Higher Mental Functioning in Operational Environments, AGARD, Paris, 1976.

14. Roscoe, A. H. Pilot workload during steep gradient approach, Conference Proceedings No 212 Aircraft operational experience and its impact on safety and survivability, AGARD, Paris, 1976.

15. Roscoe, A. H. Heart rate monitoring of pilots during steep gradient approaches, Aviation Space Environ Med 46, pp. 1410-1415, 1975.

16. Ellis G. Pilot opinion measures, Assessing pilot workload, A. H. Roscoe, ed., ACARDograph AG-233, AGARD Paris 1978.

17. Roman J. A. and Lamb. L. E. Electrocoodiograph in flight, Aerospace Med 33, pp. 527-544, 1962.

18. Rowen, B. Biomedical monitoring of the X-15 program Air Force Flight Test Centre Edward AFB, TN - 61-4, 1961.

19. Billings, C. E. et. al. Physiological cost of piloting rotary wing, Aerospace Med 41, pp. 250-258, 1970.

20. Pilot performance and heart rate during in-flight use of a compact instrument display, FAA Office of Aviation Medicine, Report No FAA-AM-75-12, Washington DC, 1975.

21. Pilot rating techniques for the estimation and evaluation of handling qualities, AFFDL-TR--68-76 Wright Patterson AFB OH, 1968.

22. Nicholson, A. N. et. al. Activity of the nervous system during the letdown, approach and landing, A story of short duration high workload, Aerospace Med 41, pp. 436-446, 1970.

23. McGregor, D. M. Lead discussion, Conference Proceedings No 106.19 Handling Qualities Criterior, AGARD, Paris, 1971.

BRAIN WAVES AND THE ENHANCEMENT OF PILOT PERFORMANCE

by

G. H. LAWRENCE, Ph.D.
Office of Naval Research
Physiology Program (Code 441)
800 North Quincy Street - Room 433
Arlington, Virginia 22217

The use of brainwaves (EEG) for the enhancement of the performance of aircraft pilots is an idea which requires, for its development, the integration of two previously independent lines of research endeavor: human performance assessment and central nervous system neurophysiology. A human performance research paradigm specifically relevant to the study of pilot performance, in the context of which the use of brain waves may feasibly be studied, will be discussed later. Attention is now directed to the state of the art of brain wave research and brain-behavior relationships, specifically those aspects which are considered to be feasibly and usefully applicable for potential use in simulated aircraft crew stations or eventually in a real-world environment.

BASIC RESEARCH

Two basic types of paradigms have been employed in studies of brain waves and performance. In the first case, spontaneous, ongoing EEG is monitored, and frequency and amplitude for some time period (usually preceeding, during, and/or succeeding some experimental treatment) are related to various aspects of performance. Usually this sort of treatment has included use of an intervening variable called, most frequently, activation or arousal; Davies and Parasuraman (1977) have identified four separate types of experimentation within this rubric. The first type of study "has attempted to discover whether decrements in (signal) detection rate, or sometimes increments in detection latency, are paralleled by corresponding changes in one or more pyschophysiological measures. . ." Another approach has attempted to identify psychophysiological (not only EEG, of course) processes and events which discriminate between periods preceding successful as opposed to unsuccessful attempts to detect a signal. A third has involved varying environmental parameters presumed to affect arousal, thereby causing an ultimate effect upon performance if in fact arousal and performance are related. Frequently these studies have led to the observation of performance changes without concommitant variation in physiological measures of arousal; as Davies and Parasuraman have put it, ". . . a dissociation of performance indices and physiological measures occurs." A fourth approach has attempted to predict individual differences in level and quality of performance from baseline scores on physiological measures, through the so-called arousal hypothesis of vigilance. Generally, this hypothesis posits an inverted U relationship between arousal and performance; at low levels of arousal errors of omission (e.g., missed detections of targets) occur, and at high levels the well known detrimental performance effects of stress and high anxiety are seen. Arousal is measured independently, usually via physiological events.

From his 1970 review, however, O'Hanlon is led to the conclusion (quoted by Davies and Parasuraman 1977) that "No reliable physiological index of alertness has been accepted, although several promising ones have been proposed. No physiological variables have been found that are as sensitive to task and environmental effects as is performance. No underlying process has been so clearly defined as to permit rational control of cerebral vigilance." Davies and Parasuraman agree with his discourgaging assessment, and suggest that methodological deficiencies result both from inadequate performance measurement and from the lack of an extant task taxonomy to make sense of the enormous number of experimental situations which have been used. Beck in his 1975 paper considers that brain waves cannot feasibily be brought under stimulus control, and therefore does not include EEG studies (as constrasted with evoked potential studies) in his review.

Nevertheless, a few semi-consistencies have been observed across experiments and laboratories, although virtually no general conclusion can be put forward which does not admit of some exception or can be considered invulnerable to challenge. Certainly decrements in detection performance, rate, and or latency are usually seen to be accompanied by EEG changes as fatigue develops over the course of lengthy experimental sessions, and it is clear that the probability of failure to respond to transitory signals altogether increases considerably under conditions of lowered arousal - i.e., when the subject is bored or sleepy. Further, several studies have indicated that individuals with greater baseline/ability for GSR (though not, so far, for EEG) do better indetection situations and generally are better able to maintain a state of vigilance. Variation in the complexity of visual stimuli, memory task requirements, and differential hemispheric activation via varying stimulus modalities have all been shown to affect EEG measures in relatively stable ways. Brain waves have been reliably shown to vary with behavioral sleep events in a relatively stable manner, and consistently over the time course of a normal nights sleep, and thus probably reflect variation in arousal (albeit at the very low end of the scale). Gales (1977) points out that ". . . high alpha and beta frequencies are more sensitive to discrete changes in stimulation than are lower alpha frequencies and theta activities. . ." indicating that the relationship of EEG events to arousal is more easily studied in alert states. Gales' statement is made in summary, and follows a passage in his paper where he acknowledges that arousal is not a unitary state which has straight-forward and systematic relationships with measures of behavior or of subjective report. He goes on to say that changes in theta and lower ranges of alpha reflect other, presumably non-task-relevant, effects.

Other attempts to relate ongoing EEG experimentally to vigilance, or to other behavior, especially under conditions of fatigue have been made. Consistent with data from other situations, it is typically found that changes in fatigue and arousal can be inferred from brain wave activity (power shifts from higher toward lower frequencies, i.e., from beta toward theta and delta).

Prepared for the Environmental Physiology Program, Office of Naval Research, USA

The work of O'Hanlon and Beatty (1977) supports the general form of the arousal hypothesis of vigilance, showing that percentage of alpha and theta increases and that beta decreases were related to variation in performance on a simulated radar watching task. As Beatty and O'Hanlon point out, alpha may either increase or decrease with arousal; concurrent changes in theta and beta must be taken into account (i.e., are frequencies increasing or decreasing?) before sense can be made of the variation in alpha.

Finally, there is some indication (e.g., Dimond 1977) that hemispheric differences may relate to arousal and vigilance performance. Split brain studies suggest that the two hemispheres may have different vigilance systems. Perhaps, as suggested by Jerison (1977), the left hemisphere deals with selective attention and the right with continuous attention.

It is probably a fair statement that there is not much promise of new and exciting use of ongoing EEG for the enhancement of pilot performance at the moment. There appears to be sufficient consistency in the literature so that some confidence may be felt in the use of changes in brain wave power across frequencies to infer rather general state changes. One can tell when a subject is getting drowsy, has gone to sleep (and brief bursts of sleep frequently appear under sustained task performance requirements, especially when some degree of sleep deprivation exists), or to a lesser degree of certainty, is simply inattentive. This kind of information is not without interest and use, but its lack of information specificity and very low data rate lead to the conclusion that the instrumentation and data processing requirements to collect and act upon it would not likely pay off in greatly enhanced performance, though its potential for monitoring organismic state is obvious.

The event-related potential (ERP) is another matter. The ERP is an EEG response evoked by a specified stimulus and usually averaged over a group of trials. A series of positive and negtative deflections is observed, usually conceptually and empirically divided into two categories. The earlier components, those occurring in the first 100 ms or so subsequent to the stimulus, are referred to as exogenous—they reflect characteristics intrinsic to the stimulus event itself, such as loudness, brightness, intensity, or other psychophysical attributes. This activity is considered to represent the processing of sensory information. The later components, up to perhaps 600 ms beyond the stimulus, are considered to be endogenous, reflecting cognitive processes and attributes of the stimulus deriving not from its physical properties but rather from its task-relevant context (e.g., whether it is to be counted or ignored, its suprisingness, its information value, etc.). It is these latter components, reflecting as they seem to aspects of performance potentially applicable to cockpit or crew station situations which are of primary interest. The following discussion of these later ERP components and their studied relationships is largely derived from comprehensive and thorough reviews by Donchin et al (1977) and Beck (1975), and to a lesser degree from the recent chapter by John and Schwartz (1978). The catalog of endogenous components offered by Donchin et al (1977) is worth presenting in full:

> N200. This component is elicited whenever a rare or unexpected event occurs. It is of particular interest because it can be elicited by stimuli that are in the periphery of the subject's attention. Unlike the other endogenous components it appears sensitive to the modality of the stimulus. The positive-going return of this component is sometimes labeled P3a.
>
> P300. This robust endogenous component is reliably recorded in association with task relevant, rare stimuli. We apply this label to a component whose latency may range from 275 to at least 600 msec. It is characterized by its scalp distribution as it tends to be larger in the central and parietal electrodes. It is further characterized by a very specific response to experimental manipulations.
>
> Slow Wave. This is a slow potential shift that, as far as is known, is affected by the same variables that affect P300, except that it has a different scalp distribution. Whereas P300 appears largely as a positive-going potential peaking on the parietal scalp, the Slow Wave is positive-going at the parietal electrodes and negative-going in frontal electrodes. As the Slow Wave is so closely associated with P300, it will not be discussed further in this report.
>
> The Contingent Negative Variation (CNV). This term is normally used to describe the slow negative shift of potential that occurs the warned foreperiod preceding a motor or mental task. It begins approximately 400 msec after the warning stimulus and, normally, terminates after the imperative stimulus, that is, the stimulus demanding a response or decision by the subject.
>
> The Readiness Potential (RP). As the CNV, this is an event-preceding negative shift. It is distinct in the sense that it appears prior to self-paced voluntary responses. Its occurrence is independent of the presence of an eliciting, or command, stimulus.

As John and Schwartz (1978) stated, these endogenous components have been studied in connection with arousal, attention, selective attention, emotional valence, assessment of novelty, time estimation, uncertainty, detection of targets, differential identification of stimuli indepedent of size and shape, and the semantic classification of linguistic symbols.

Some of the more potentially useful and applicable findings that have occurred with some consistency are:

o it is clear that ERP component amplitude is related to attention For example, P300 (whose latency actually varies from 250 to 600 ms) is rather significantly amplified by the perception of a meaningful or suprising stimulus (or by the absence of a stimulus, when one is expected); in other words, by the resolution of uncertainty. This basic finding has been observed in situations involving reaction time, signal detection, signal confirmation, pattern completion, motor set, as well as other experimental paradigms. The finding that P300 occurrence follows omission of an expected stimulus seems especially interesting, demonstrating clearly that this component reflects cognitive rather than sensory processing. The potentials evoked by missing stimuli are indistinguishable in confirmation, from those evoked when sensory stimuli are in fact presented. Results of studies utilizing the omitted stimulus paradigm allow the inference that an internal model of sequences of stimulus events is formed, and that the P300 is evidence of a mismatch between this model and the observed (non) event which unexpectedly does not occur.

o in general, the P300 is enhanced only when stimulus information is being actively processed and is uniquely associated with the occurrence of a signal and its correct detection. (Beck 1975). It occurs subsequent to stimuli in any sensory modality. Evidence exists that amplitudes and latencies vary over the scalp, and appear to interact with different task analysis requirements for cognitive processing.

o there is some indication that auditory ERP's vary interhemisperically as a result of a linguistic task; left-side responses are larger when a task requires linguistic analysis but not when the same stimuli are compared non-linguistically.

o John's (1978) review of P300 variation relating to semantics and logic, and varying over the scalp in amplitude and latency, lead him to the conclusion that:

> these articles (e.g., Thatcher, R. W. 1976) provide compelling evidence that the features of the ERP are influenced dramatically by subtle semantic relations such as whether stimuli are synonymous or antonymous, whether they are bilingual equivalents or not, or whether they are logically true or not. The changing latency and/or anatomical locus of the LPC differences, observed as these information parameters shift within an overall match-mismatch paradigm of otherwise constant design, provide perhaps the most convincing evidence that these ERP features relate to cognitive mental processes rather than the nonspecific or physical factors in these experiments.

The CNV has also proved to be a fertile source of research efforts attempting to relate aspects of its occurrence to stimulus attributes. The slow negative shift of brain potential occurring between a warning and an action stimulus which is referred to as CNV has attracted the attention and interest of a number of investigators, resulting in a body of literature summarized by Beck (1975) into groups of studies which interpret the CNV as reflecting expectancy, motivation, conation, or attention. Beck asserts the overall implications that (a) the CNV is not a single process, and (b) that its nature and cerebral topography are dependent upon the state of the organism and the task imposed. He points out that CNV magnitude has been seen to relate to the uncertainty, intensity, and amount of information in the action stimulus, the interstimulus interval, concentration, and anxiety.

Rosenweig and Leiman (1968) provide a useful summary of the major research approaches to the study of brain-behavior relationships:

> Interest in the relation between electrical indicants of neural function and dimensions of behavior has taken several forms. First, considerable research has been directed toward the examination of the neural representation of sensory inputs, particulary the relationship between neural coding and generalizations derived from the data of psychophysical experiments. A detailed specification of the transfer operations at relay nuclei in particular sensory systems has been noted by several investigators. Second, various investigators have sought to relate electrical parameters of neural functioning to more elaborated behavioral phenomena--dimensions such as attention and learning. This type of experimental inquiry has been direct to analyses of mechanisms of integration and to relating regional electrophysiological differences to particular attributes of behavior. Experimental stratagems have been diverse, although most studies can be considered as either (a) exhaustive analysis of the synaptic mechanisms of discrete functional systems which may form a basis for the analysis of more complex operations mediated by these systems; or (b) determination of the behavioral correlates of particular intrinsic wave processes, e.g., alpha rhythm or hippocampal theta rhythm; or (c) assessments of the induced electrographic states produced by special behavioral manipulations e.g., evoked activity and reaction time. (pp. 69-70)

EEG has several attributes which result in its popularity for use in situations where monitoring is required; it occurs continuously and spontaneously, and it is readily available via relatively inexpensive and simple, instrumentation. Spontaneous, ongoing EEG can be automatically recorded and analysed with relative ease, and has served as an indicator of many disparate varieties of states ranging from the clincial (e.g., death, schizophrenia, anxiety) to performance (e.g., arousal, perception, processing efficiency). As Mirsky (1969) has pointed out:

> . . . large masses of data may be processed quickly; the data so obtained may be more easily subjected to statistical evaluation. Methods range from total integrated energy (more properly, voltage) through base-line crossing, frequency analysis, and analysis of slope changes into amplitude measures. (p. 325)

Evoked electrical potentials differ from ongoing EEG by occurring in close temporal proximity to the stimuli by which they are elicited, by their relative consistency of shape, and (usually) by much smaller amplitudes than background brain wave activity.

Sem-Jacobsen (1971) has monitored EEG (among other physiological indices) from pilots under stress in real-life missions, and has shown that previous quality of pilot performance can be used to predict occurrence of delta-theta (2 - 8 hz) activity under stress. Sem-Jacobsen attributes many aircraft accidents to pilot task overload which elicits a freezing under conditions of extreme stress, reflected in a showing and eventual flattening of brain wave activity.

A POSSIBLE SYNTHESIS

A recent paper by Cooper et al (1977) presents an interesting demonstration of cortical potentials, large enough to stand out from background EEG and be recognized easily without averaging, which occur as a result of changes in a visual display consisting of a video recording of a model landscape, comprising an aerial view of countryside with residential buildings, fields, hedgerows and roads in the foreground, and with mountains and sky in the distant background. Various vehicles (vans, trucks, or cars) were introduced and moved through the scene at varuous infrequent, unpredictable times. The task of the subjects was to (a) detect and then (b) identify by type, each vehicle which entered the display screen. Typically, about a second before the detection response was made, S's eye movements indicated a sudden shift to the interesting for the present consideration) "a large, discrete, well-defined, positive-going potential (occurred) in the EEG." This brain the range of this latency varied from 0.1 to 2.0 seconds, which rules out a simple analogy to the P300 paradigm with detection the eliciting stimulus. There was no such wave, incidentally, associated with the other half of S's task, the identification of the type of vehicle.

As the authors point out, since

> . . . it appears that the positivity occurs when the observer sees one of the class of events that he has been told to detect.
>
> Both the role of these potentials, their distribution, and their positive polarity suggest that they might have common origins with the P300 component of the cortical evoked potential which occurs characteristically during discrimation and decision-making tasks.

A less well-defined brain event observed in this experiment was

> a slow increase of negativity that started before the detection positivity . . . (which) can start to increase while the eyes are scanning other parts of the display, and it decreases after completion of the motor tasks indicating detection and recognition.

Although the possible relationship of this observation to the CNV previously described seems a tempting area for speculation, the authors refrain from its pursuit. It is, however, a most appropriate area for research, and an elucidation of the relationship of these phenomena with P300 and CNV might open the door to the application of the rather considerable body of knowledge which has already been gathered about these brain wave events in laboratory settings to situations much more veridical than the typical laboratory problem now utilized.

THE WORKLOAD PARADIGM

Perhaps more progress has been made toward the utilization of brain wave information for the enhancement of pilot performance in the area of monitoring and assessment of workload than in any other area. This concept has been pursued effectively by Donchin and his colleagues (Wickens, Isreal, and Donchin 1977, e.g.), in the context of a large-scale effort which is centered in Donchin's Cognitive Psychophysiology Laboratory at the University of Illinois and is aimed at the development of very closely coupled man-computer systems. Wickens (1978) provides a brief but thorough and clear description of the purposes of the workload measure and several hints as to the potential operational significance of information derived therefrom:

> In the on line adaptive context, it may be asserted that an intelligent computer-based adaptive system, in order to optimally deploy the resources of human and computer, should be provided with a real-time, updatable estimate or model of the state (availability and allocation) of the operator's attentional resources, so that adaptive procedures may be initiated. The characteristics of this state estimate and their potential use to the adaptive decision maker are as follows.
>
> (1) The available resources must be sufficient to meet the demands imposed by all tasks which challenge the operator at any time: the characteristic of task workload or reserve capacity. If the momentary workload demands become sufficiently great, adaptive aiding procedures can be implemented to temporarily unburden the human operator. These may take the form of initiating alternate strategies of computer processing or information display, implementing automatic control systems, or of calling for extra manual assistance where such is available.
>
> (2) Even if the resources are adequate, the attention must be allocated *properly* to the critical tasks, display or sources of information, so that important sources are not ignored: the characteristic of attention allocation. The distinction between

> workload and allocation is crucial. It is self-evident that adequate capacity inadequately deployed may lead to non-optimal performance. Thus if resources are allocated incorrectly along channels of incoming information or, if critical sources of information are being ignored, warning or cueing signals might be provided to redirect the proper distribution of attention as determined by some preset establishment of priorities. Alternatively, system characteristics might be alerted with derived knowledge of the distribution of operator attention. For example, computer resources might be allocated to monitor for, or process signals along, information channels that are inferred to be non-attended.

The research basic paradigm which is at present envisioned for the evaluation of workload measures and their incorporation into a computer-aided aircraft control system involves an operator working on one or more tasks which can be varied in complexity and difficulty. A flash of light, or similarly transient auditory stimulus, is introduced; this stimulus must be counted or otherwise used in the performance of a secondary (or teritary) task, and the latencies and amplitudes of the P300's elicited by these so-called probes (i.e., tests of attention, reserve processing capacity, etc.) are used as indicators of operator workload; the controlling computer then allocates task responsibility between man and machine primarily on the basis of this information, thereby optimizing overall system performance. Too little work for the operator will result in boredom and performance deterioration, so it may be wise to require of the man some work which theoretically is most efficiently and competently undertaken by machine. Too much work for the operator, particularly under especially stressful conditions (e.g., landing in bad weather, combat, maneuver in crowded airspace) may result in dangerous overload and, again, deterioration of performance; so too there are times when a function which is theoretically best handled by a human should (or even must) be assigned to the machine. The overall guiding principal is, obviously, to use knowledge of the operator's reserve processing capacity and level of performance in the light of current task demands.

Wickens (1978) asks,

> ". . . what represents an appropriate control, non-adaptive system against which to compare adaptive system performance? Should this control be one that functions with maximum aiding or minimum, or an intermediate level? Should it perform under best-case or worse-case environments? The latter refer to those conditions that are not typical, of real world operations, occurring only very infrequently, but with potentially disastrous consequences? What performance index should be used to evaluate and compare such systems?
>
> As a specific hypothetical example, consider the following: An adaptive multi-loop control system is developed which will shift an axis to automatic autopilot control, when concurrent task workload is inferred to excede some criterion. A tracking performance index of this system can clearly be measured. However, against what should it be compared? The performance of a maximum-aiding (all autopilot) system--to which it will almost inevitably be inferior, or to that of a minimum aiding (all manual) system to which it will probably be superior? To complicate matters, this superiority relationship might be reversed by introducing "worse case" environments: e.g., levels of disturbance which the autopilot is incapable of handling, or autopilot system failures which, it has been argued, are less easily detected under autopilot that in-the-loop situations (Young 1969). Under these circumstances, what arbitrary costs should be assigned to latencies in failure detection, as opposed to errors in tracking performance, when deriving an ultimate performance index? These again, are questions whose answers may lie beyond the scope of the conference, but which inevitably arise when adaptive uses of workload measures are considered.

Some of the experiments which have provided the basis for the development of his paradigm, and which for the most part emanate from the Cognitive Psychophysiology Laboratory at the University of Illinois, will now be briefly described. As Donchin (1976) has stated in general,

> specific components of the ERP's have been shown to be manifestations of such cognitive events as the preparation to respond, the preparation to intake and process information, the registration of a surprising event, or the processing of task-relevant information. Our studies during the past two years have indicated that the ERP is a particularly valuable source of information about the operator, information which either is not otherwise available, or which can only be obtained through potentially disruptive direct questioning. Briefly stated, the ERP appears to be a sensitive index of the strategies adopted by individuals to cope with their assigned tasks, the manner in which they allocate their resources and distribute their information processing capacities.

EXPECTANCIES

A good example of a large group of experiments on the way in which S's mental set or expectations match the information conveyed by the stimulus is Squires et al (1976), which demonstrates changes in P300 amplitude and conformation change "as a function of the prior probability of the stimulus and the specific sequence of the preceding stimuli." In other words, for any given level of frequency of an event (the intrinsic probability of the event) the suprisingness of any single occurrence of it is significantly affected by the value of its immediate predecessors. Under conditions of relatively low workload a previous sequence of stimuli up to about five is taken into account; when the workload is high, the number of previous stimuli which are apparently taken into account by S decreases. This slope, which can be reliably shown to be a function of task demands, is interpreted as an index of the reserve processing capacity of S's short-term memory buffer, and therefore is considered a useful means for the assessment of workload (Duncan-Johnson and Donchin 1977). It is interesting to note that this effect has been demonstrated experimentally to hold over a range of intrinsic probabilities of .10 to .90 (Duncan-Johnson and Donchin 1977), that the interval between probes is not of critical importance (Donchin, McCarthy and Kutas 1977), and that there are some modality differences (auditory vs visual) in (1) the way workload affects changes in P300 amplitude and (2) in the operation of the sequential effect model described above. Specifically non-target visual stimuli (i.e., visual stimuli to which S need not respond) do not, apparently follow the model for P300 amplitude (Squires et al 1977). Incidentally, consideration of the use of the sequence effect described above should be tempered by the realization that this effect disappears when the S knows the intrinsic probabilities in the situation. Perhaps an implication of this, and it might be a useful one, would be that in a situation of known probability parameters, a rare and important stimulus which occurred several times in a row would not rapidly become habituated. The boundaries of this situation, i.e., the length of the epoch (in which the postulated rare stimulus occurred frequently) which would be necessary before S decided that this event had now become frequent, at least temporarily (as opposed to perceiving, simply an unusual frequency of a truly rare stimulus), might be interesting as an experiment. How rapidly does a rare or unusual stimulus habituate and what factors affect the rate?

In general, the P300 relates to a stimulus through changes in latency-which reflect the speed with which the stimulus is recognized-and amplitude, which reflect the informational value and task relevance of the stimulus. Informational value, or feedback, is usually operationally represented in these experiments by stimulus rarity; task relevance, through instructional set. In general, greater task relevance increases amplitude and greater attentional demands (e.g., greater display complexity) increase latency. The stimulus must have at least some task relevance in order to elicit a P300 at all. If it is an unusual or rare (low probability) stimulus, and therefore a surprising stimulus (the degree of which may be affected by preceding stimulus patterns, knowledge of intrinsic probabilities, and, no doubt, other circumstances), it will elicit a relatively large P300.

intelligence, motivation, e.g. - any or all) which are to be allocated among various demands at any given time. A computer-based performance enhancing system should monitor the current resource allocation, the additional leftover utilizable capacity, and be able to program optimal sharing of task responsibility between itself and the human operator. Usually the primary task involves tracking, and its difficulty can be readily manipulated (e.g., by varying the number of dimensions); the secondary task involves, usually, counting the less frequent member of a pair of tones which vary in pitch.

Experiments utilizing this paradigm "indicate that while first order ERP's are relatively insensitive to momentary fluctuations in tracking difficulty, they clearly discriminate between low levels of tracking demand (no tracking vs. one dimensional tracking). Higher levels of demand (one vs. two dimensions) are differentiated by the extent of sequential processing of the stimulus series, a measure that is similarly revealed in the ERP's." (Donchin 1976). Taken together, these data provide the basis for a tentative measure of workload (and, more importantly, the obverse: reserve resource capacity) based upon amplitude of P300. This model has been useful, for example, in experiments (1) showing that a secondary task requiring an occasional button-press interferes with performance on a (difficult) tracking task and (2) studying the effect when a third task is added to ongoing primary and secondary tasks. In the former case, if whatever action is taken as a result of the button-push is taken instead when P300 occurs, hypothesizes Donchin (1976), performance on the tracking task would show less interference effect. In the latter case, the slope of the sequential effect (i.e., the number of previous stimuli which affect P300 amplitude to the auditory probes comprising the secondary task) determine the likelihood that the third task can be introduced without deterioration of performance on tasks I and II. It should be noted that performance on tasks I and II are not enhanced in this model (though it is likely that information derived therefrom could be used to optimize all-task performance by changing parameters when P300 indicates low processing capacity); it is the overall system effectiveness which benefits. In other words, the third task could be introduced only when the operator is cognitively ready and able to handle it.

Another experimental paradigm of interest is the comparison of (1) the relationship of P300 latency to reaction time in speed-demand situations with (2) this relationship under an accuracy-demand instructional set. Under speed conditions response generally precedes P300; under accuracy conditions, reaction follows P300. Experimental results (Kutas, McCarthy and Donchin 1977) show that there is a high probability that if an error has been made under speed conditions, reaction time is less P300 latency in a paradigm where S is required to count the rare stimuli, and that making use of this knowledge can enhance performance under the speed condition to the level attained under the accuracy condition - at no sacrifice in speed.

There are other experimental methodological ingenuities which appear interestingly relevant. The Cooper et al study which seems to elicit brain events which look like P300 in quasi-veridical settings has already been discussed, and would appear to offer the means to take into account simultaneously information to be gained from on-going monitoring of background EEG with naturally-occurring ERP's. Another idea which may be promising is Regan's (1977) work on steady-state ERP's. Here a transient change of intensity or some other important parameter of a sensory stimulus is repeated to elicit a series of ERP's which, through Fourier transform analysis, can be used to monitor cognitive response to the stimulus. Finally, the finding by Kutas (dissertation in progress) that hemispheric CNV amplitude indicates intended choice of hand with which a response is to be made (in a left-right discrimination task) might be eventually put to good operational purpose.

PROBLEMS AND APPLICATIONS

It may be useful at this point to group the sorts of pilot problems which may lend themselves to brain wave enhancement, and to speculate briefly as to some of the ways in which the experimental findings described above may, eventually, apply.

WORKLOAD ALLOCATION

Workload allocation reflecting accurate continuous assessment of operator state and processing capacity probably would function for pilot performance enhancement most frequently in normal flight, to provide the earlier mentioned theoretically optimum mix between human and machine control of aircraft function. Donchin (1976) describes an intended experiment which will test the workload allocation model described previously and which could serve as a prototype for an operational aircraft operation:

> The function of an autopilot is to replace the human as an active element in the flight control loop. The reduction in pilot work-load gained when the aircraft functions in an autopilot mode is however purchased at a cost. The cost is represented by the increasing likelihood of a malfunction in the control system with the greater number of mechanical elements in the loop, the loss of flexibility of the system in responding to unexpected environmental inputs, and the decrease in speed and accuracy with which the pilot can resume command if such events occur (e.g., the appearance of an input requiring an evasive maneuver or implementation of some other control strategy).
>
> Because there is a cost associated with autopilot control based upon actuarial data of component reliability and on the likelihood of external event occurrence, such control should not be in effect al all times, but only when the external (non-control) demands of the operator reach a sufficiently high level that the cost of remaining <u>in the loop</u>, as determined by excessive work-load, exceeds that of autopilot function.
>
> This experiment will explore the implementation of an ERP-based adaptive algorithm that will shift control from human to autopilot, according to inferences based upon total processing demand and the distribution of processing resources.
>
> Subjects will fly a predetermined simulator flight path, perturbed by an intermediate level of turbulence. External demands will be imposed in the form of a discrete decision making task, stimulating target identification and classification. Subjects will be instructed to monitor a display for discrete items of alphanumeric information and identify and classify this information into one of four categories via a manual response.
>
> The subject's cognitive state will be monitored by three kinds of ERP inducing stimuli. The requirement to process a Bernoulli series will generate a measure of overall work-load. ERPs elicited by the appearance of the peripheral targets will be analyzed to infer the degree of processing of these targets. Finally, probe stimuli delivered along the channels conveying flight path information will elicit ERPs that will be employed to make inferences concerning processing or neglect of the axes of tracking control.
>
> An adaptive algorithm will monitor the ERPs and make the decision to implement autopilot aiding according to the following decision rule. If work-load is inferred to be high <u>and</u> and tracking axes neglected, the autopilot will be implemented. Otherwise the pilot will remain in the control loop. Once the pilot is out of the loop, the decision to de-activate the autopilot will be made if the level of work-load and peripheral target frequency both drop below predetermined criteria. Performance of the adaptive system will be based upon a joint measure of target identification performance (speed and accuracy), and tracking error, integrated over both the manually controlled and autopilot deviations. From this index will be subtracted a fixed cost per minute of the time spent in the autopilot mode. This performance index will be compared with that achieved in a regular non-adaptive session of the same task in which naturally the autopilot cost term will be zero.

Computer-controlled workload allocation could also function in an important beneficial manner under acutely demanding and stressful conditions, applying the same basic algirithm in a slightly different way.

WARNINGS

The use of brain waves for the automated enhancement of warning effectiveness could occur in two ways. A computer could sense some deficit in an operator's state of being, or potential deficit (anticipating a crisis), or the computer could observe, perhaps, a lack of attention to a warning display or other performance (as opposed to state) deficit, and take action to further stimulate the human operator. In the former general circumstance the system monitor would make inferences about operator state, probably from a set of physiological information channels; in the latter, it would make inferences about observed deficits in operator performance, probably from assessment of ERP's or their lack, in response to warning signal displays treated functionally for this purpose as probes. The potential for use of brain wave indicators of dangerous operator state or behavior is doubtless apparent: presence of theta can predict drowsiness and deterioration of performance, and of course the sleeping state can be readily detected. The detection of undesirable levels of arousal (inappropriately high or low) or emotional states can probably be enhanced through other physiological or behavioral channels. Attention to a display could be assessed via Regan's (1977) steady-state ERP technique mentioned earlier. Donchin (1976) holds out the promise of use of ERP's to distinguish non-response to a warning resulting from a purposeful decision to ignore it, from accidental non-recognition; this way the computer-based system can refrain from repetition or intensification of information which the operator has already processed and to which he presumably will respond.

SENSING EVENTS AND SAVING TIME

The computer could also certainly determine the occurrence of an event like target acquisition (the Cooper et al paradigm, e.g.) from occulography, pupillometry, and brain waves more rapidly and less disruptively than this information can be made known by a human observer employing a gross skeletal response such as pushing a button. In some cases it may even provide a more accurate judgement of the task-relevant event than could the button push (i.e., the P300 latency - reaction time relationship under speed conditions described earlier). In any event, there are circumstances wherein a few milleseconds could provide an important advantage. Use of information about the laterality of the readiness potential to (a) infer the imminence of a motor response, (b) the hand with which it will be made, and (c) to execute the implied command (e.g., fire a weapon, change course and/or speed, transmit a message) or (d) to bring a control device to the desired hand (to avoid reaching), might significantly affect performance efficiency - especially if the small single savings in time and effort were to be accumulated over rapidly successive events in a continuing, recycling context of swift decision making and response.

FARTHER ALONG

At some point in the development of more highly interactive, closely-coupled man-machine interfaces, a serious effort should be made to develop the capacity for real time thought commands. In this mode the computer would sense specific wishes and needs (and evaluations of the adequacy of its own moment-to-moment performance in meeting these needs) on the part of the operator. Ultimately the ability to infer, accurately, sequential chunks of complex information would be needed, utilizing electrical representations of verbal or non-verbal cognitive activity. Chapman's efforts (1977) to locate ERP's related to specific words in multi-dimensional semantic space via the semantic differential technique may be a promising step toward this end.

A more proximal goal would be the development of machine ability to sense such general intangibles as operator uncertainty (and therefore the need for more information, or at least a need to maintain decision options), and approval or disapproval. Sensing approval or disapproval (for want of a better descriptor) would provide instantaneous qualitative feedback to the machine somewhat in the way "warm-cold" feedback is provided to the blind searcher in children's games - or, perhaps more appropriately, in the way varying intensities of temperature guide a missile toward a heat source. Ability to assess these variables continuously and sensitively could provide the basis for very fine control of machine by man, perhaps even allowing the creation of an artifically intelligent servomechanism so closely responsive in real time to its operator's cognitions that it could serve virtually as a functional extension of the operator's own nervous system.

The ultimate aim for this type of man-machine system in general, and a goal at least as well suited to enhancement of aircraft pilot performance as to any other military application, is the utilization of the human operator for those purposes for which he is uniquely qualified: as a complex pattern recognizer and decision maker - the pure strategist or tactician. The computer would, in real time and functionally as if an organic part of the operator, undertake such activities as storing, organizing, and retrieving data base information as it is acquired or needed, performing other data handling functions, and carrying out decisions once it can accurately determine that they have been made.

CURRENT PROBLEMS AND NEEDED RESEARCH

Aside from the obvious, which is to increase the certainty with which inferences from P300 can be made and to refine the methodology of making use of them, it is possible to describe several areas where current research techniques need to be made more powerful, and new methods which have been identified and require further development.

Virtually total reliance on P300 for access to cognitive events is too limiting, and there are several ways in which more brain information might be made available. One line of attack into this problem area would be to seek to understand the events underlying other components. Vidal (1977), for example, has made potentially interesting use of some of the early exogenous components to guide a cursor. Also, there is some indication that P300 bandwidth might be increased by independent probing of isolated sectors of each retinal field. Further, Donchin (1976) has suggested some conditions under which N100 (an exogenous component) and N200 might yield operationally useful information on attentional, perceptual, and processing events. Another line of investigation would be to open up new sources of information. For example, using multiple arrays of electrodes, latency, and amplitude differences arising from different sites might reflect distinct cognitive activities. Further, regional variations in latency and amplitude of the same component might be related in stable ways to various cognitive activities. Development of functions representing, say, ratios of P300 amplitudes at various locations combined with concurrent sets of latency differences, or other secondary treatments of multiple recordings of the same event, might yield fine discriminations among processing stages or other relevant aspects. Also, the development of magnetoencephalography would allow access to subsurface activity (as well as providing physically non-impinging sources).

Even better and more reliable single-trial identification of brain wave events of interest is needed. While the development of Donchin and his colleagues (e.g., Squire & Donchin 1976) of such a capability, using a sliding template and stepwise discrimination analysis has made feasible the real time use of ERP's for vehicle control, much more needs to be known about such issues as the stability of individual templates for recognition of ERP's, optimum strategies for updating these templates, and, of course, increasing the accuracy with which such recognitions are made. Present capabilities seem remarkably good, but if life-or-death actions are to be taken on the basis of them, either accuracy must be improved or a fail-safe procedures developed. For allocation of workload under routine or even somewhat demanding conditions present levels of accuracy seem adequate.

REFERENCES

Beck, E. C., "Electrophysiology and Behavior." Ann. Rev. Psychol. 1975, 26. Palo Alto: Annual Reviews Inc.

Chapman, R. M., Blagden, H. R., Chapman, J. A., and McCrary, J. W., "Semantic Meaning of Words and Averaged Evoked Potentials." In: LANGUAGES AND HEMISPHERIC SPECIALIZATION IN MAN: CEREBRAL ERP's. Prog. Clin Newrophysio., 3, (Ed.) J. E. Desmedt (Karger:Basel, 1977)

Davies, Dr. and Parasuraman, R., "Cortical Evoked Potential and Vigilance: A Decision Theory Analysis." In VIGILANCE: THEORY, OPERATIONAL PERFORMANCE AND PHYSIOLOGICAL CORRELATES. Mackie, R. R. (Ed.) New York: Plenum 1977.

Dimond, S., "Vigilance and Split-Brain Research." In VIGILANCE: THEORY, OPERATIONAL PERFORMANCE, AND PHYSICLOGICAL CORRELATES. Mackie, R. R. (Ed.) New York: Plenum 1977.

Donchin, E. 1976. Personal Communication

Donchin, E., McCarthy, G., and Kutas, M., "Electroencephalographic Investigations of Hemispheric Specialization." In: LANGUAGE AND HEMISPHERIC SPECIALIZATION IN MAN: CEREBRAL ERP's. Prog. Clin. Neurophysio., 3, (Ed.) J. E. Desmedt (Karger: Basel 1977).

Donchin, E., Ritter, W., and McCallum, W. C., "Cognitive Psychophysiology: The Endogenous Components of the ERP." In CALLAWAY, E., Teueting, P., and Koslow, S. (Eds.) In press (Academic Press) 1978.

Duncan-Johnson, C. and Donchin, E., "On Quantifying Surprise: The Variation of Event-Related Potentials with Subjective Probability." Psychophysiology, 1977, 14, 456-467.

Gale, A., "Some EEG Correlates of Sustained Attention." IN VIGILANCE: THEORY, OPERATIONAL PERFORMANCE, AND PHYSIOLOGICAL CORRELATES. Mackie, R. R. (Ed.) New York: Plenum 1977.

Jersion, S., "Vigilance: Biology, Psychology, Theory and Practice." In VIGILANCE: THEORY, OPERATIONAL PERFORMANCE, AND PHYSIOLOGICAL CORRELATES. Mackie, R. R. (Ed.) New York: Plenum 1977.

John, E. R. and Schwartz, E. L., "The Neurophysiology of Information Processing and Cognition." An. Rev. Psychol. 1978, 29, Palo Alto: Annual Revies Inc.

Kutas, M. and Donchin, E., "Studies of Squeezing: Handedness, Responding Hand, Responded Force, and Asymmetry of Readiness Potential." Science, 1974, 186, 545-548.

Kutas, M., McCarthy, C., and Donchin, E., "Augmenting Mental Chronometry: The P300 as a Measure of Stimulus Evaluation Time." Science, 1977, 197, 792-795.

Mirsky, A. F., "Neuropsychological Bases of Schizophrenia." Ann. Rev. Psychol. 1969, 19. Palo Alto: Annual Reviews Inc.

O'Hanlon, J. F. and Beatty, J., "Concurrence of EEG and Performance Changes during a Simulated Radar Watch and Some Implications for the Arousal Theory of Vigilance." In VIGILANCE: THEORY, OPERATIONAL PERFORMANCE, AND PHYSIOLOGICAL CORRELATES. Mackie, R. R. (Ed.) New York: Plenum 1977.

Rosenzweig, W. R. and Leiman, A. A., "Brain Functions." Ann. Rev. Psychol. 1968, 18. Palo Alto: Annual Reviews Inc.

Regan, D., "Steady-State Evoked Potentials." J. Opt. Soc. Am., 1977, 67, 1475-1489.

Sem-Jacobsen, C. W., "Physiological Aspects of Aircraft Accident Investigation." Aerospace Med., 1971, 42, 199-204.

Squires, K. C. and Donchin, E., "Beyond Averaging. The Use of Discriminant Functions to Recognize Event Related Potentials Elicited by Single Auditory Stimuli." Elect. & Clin. Neurophysiol. 1976, 41, 449-459.

Squires, N. K., Donchin, E., Squires, K. C., and Grossberg, S., Bisensory Stimulation: Inferring Decision-related Processes from the P300 Component." J. Exp. Psychol. Hum. Perc. and Perf. 1977, 3, 299-315.

Squires, K. C., Wickens, C., Squires, N. C., and Donchin, E., "The Effect of Stimulus Sequence of the Waveform of the Cortical Event-related Potential." Science, 1976, 193, 1142-1146.

Wickens, C. D., Position Paper, NATO Conference on Workload. In press, 1978.

Wickens, C. D., Isreal, J., and Donchin, E., "The Event Related Cortical Potential as an Index of Task Workload." Proc: 1977 Ann. Mtng., Hum. Fac. Soc.

Vidal, J., "Real-Time Detection of Brain Events in EEG." Proc. of the I.E.E.E. 1977, 65, 633-641.

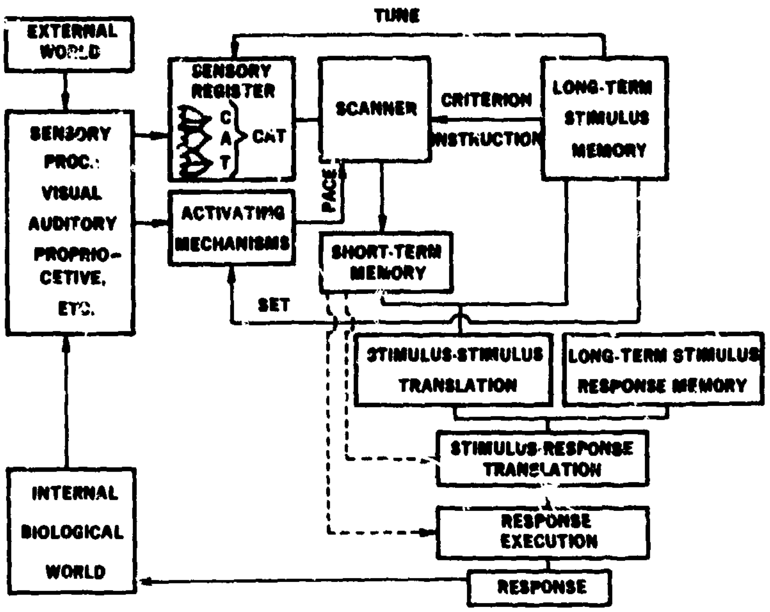


Figure 1. Diagram of Teichner's Theoretical System (after Teichner, 1974).

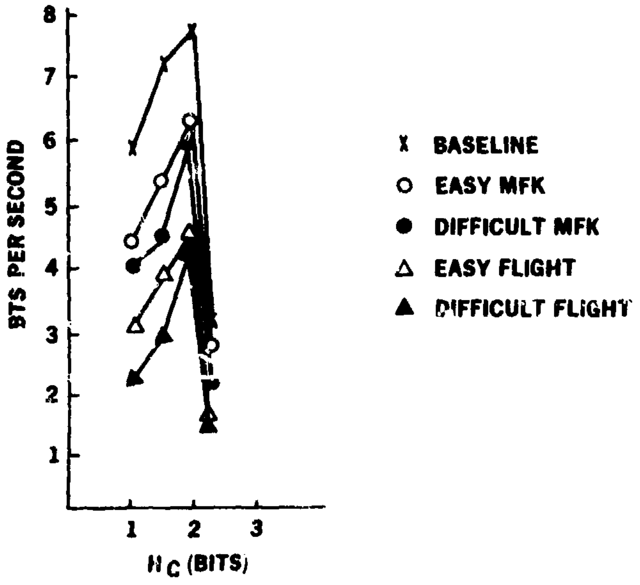


Figure 2. Average rates of effective information reduction on Sternberg task.

PUPILLOMETRIC METHODS OF WORKLOAD EVALUATION: PRESENT STATUS AND FUTURE POSSIBILITIES

by

Jackson Beatty, Ph.D.
Department of Psychology
University of Califonia at Los Angeles
Los Angeles, California 90024

INTRODUCTION

The assessment of pilot workload is a special case of the measurement of information-processing load, the aggregated demands placed upon an individual in the performance of a particular cognitive task or function. Three general approaches have been employed in the measurement of information-processing load. The first is that of subjective estimation. Subjective estimates are involved when workload is estimated from the task engineer's opinion as to the probable magnitude of processing load, an opinion that may be based on previous experience or an analytic theory. However, subjective estimates of workload by the user or participant are the most common form of workload measurement in aircraft design. Both types of subjective ratings have serious weaknesses.

The second major method of measuring processing load employs behavioral measurement. Here the notion is that the information-processing capacity of an individual is limited so that the workload imposed by one task can be estimated by the degree to which it interferes with the simultaneous execution of a secondary measurement task, such as simple reaction time or manual tracking. This method has much to recommend it over the subjective measurement techniques, particularly with respect to objectivity. But the behavioral-interference method is difficult and time-consuming to implement, and yields relatively little data for the amount of time and energy invested in testing. As a consequence, this method has been of more theoretical than applied interest.

The third major method is physiological, in which the response of the nervous system to the load imposed by an information-processing task is assessed. Momentary increases in processing load induce short-latency, short-lived increases in measures of central nervous system activation. These changes are most evident and most easily measured in the autonomic nervous system. Among the autonomic measures of activation, changes in pupillary diameter appear to be the most sensitive and accurate (Kahneman, Tursky, Shapiro & Crider, 1969).

This paper discusses the use of pupillometric measures in the evaluation of pilot workload. I begin by describing the innervation of the pupil with respect to its connections with brainstem activation systems. Modern methods for pupillometric measurement are then described. Next, a series of experiments describing pupillary response in a variety of information-processing tasks is reviewed. Finally some possibilities for the use of pupillometric methods in the measurement of pilot workload are discussed.

INNERVATION OF THE PUPIL

Pupillary diameter is determined by the relative state of contraction of the two opposing muscle groups of the iris, the sphincter and the dilator pupillae. The dilator pupillae are radially oriented bands of smooth muscle that are innervated by the sympathetic branch of the autonomic nervous system through the cervical sympathetic ganglia. The sphincter pupillae are innervated by the parasympathetic system through the ciliary ganglia, and act to close the pupil when activated. Pupillary dilation, therefore, can result from either sympathetic activation or parasympathetic inhibition. Cortical inhibition of the Edinger-Westphal nucleus, the brainstem nucleus that projects to the ciliary ganglia, has been frequently hypothesized to accompany cortical activation. Both the sympathetic and parasympathetic brainstem nuclei involved the regulation of the iris musculature are intimately connected with the reticular activating system. Indeed, pupillary measures were used to assess reticular formation functions in the pioneering work of Moruzzi and Villablanca (Moruzzi, 1972).

METHODS FOR PUPILLOMETRIC MEASUREMENT

The pupillometric measurement of information-processing load requires that accurate measures of pupillary diameter be obtained during the course of an information-processing task. Early work in this area employed photographic methods, in which infrared photographs of the pupil were obtained at the rate of 1 or 2 per second for the duration of the processing task. Pupillary diameter was determined by subsequent direct measurements taken from the enlarged image of each frame. Although accurate, the photographic method suffers from two weaknesses: it is a laborious and frustrating procedure to implement and, for this reason, it is not practical to obtain fine temporal resolution because of the resulting proliferation of photographs.

The second principal method of pupillometric measurement involves the use of a high-resolution infrared video camera and a special-purpose image processor that extracts an estimate of pupillary diameter from each frame of the video image. Originally developed under a grant from NIH, this instrument is presently manufactured by Gulf and Western Applied Science Laboratories, formerly the Whittaker Corporation. All major laboratories now involved in pupillometric research use this instrument.

In the basic video scan pupillometer the subject's head is restrained by a chin and forehead support. An infrared video camera is placed outside the subject's foveal field of vision, as is an infra-red slit-lamp illuminator. Both the illuminator and the camera are focused on one of the subject's eyes and the resulting image is sent to the image processor for extraction of pupillary diameter. This basic configuration of head rest, illuminator and camera is adequate for most experimental work. Pupillary measurements may be made over a wide range of lighting conditions, including complete darkness. The subject is free to move his gaze over a limited portion of the visual field; strict fixation is not required. Positioning of the subject in the head support is quickly accomplished. With appropriate adjustments this testing arrangement results in little subject fatigue.

For purposes requiring greater freedom of head movement than allowed under this configuration, a head-tracking pupillometer may be used. This permits recording of pupillary diameter from a seated subject with complete freedom of head movement. In this device, two video cameras are employed. The second camera is used to locate the head of the subject in three-dimensional space and by the use of servo-mechanisms direct the primary camera to the subject's pupil in that space. Although rather expensive, this head-tracking arrangement seems to perform quite reliably.

In the basic pupillometer, pupillary diameter is estimated from the video image of the eye by the following method: Each raster line of the image is first scanned for sharp light/dark contrast points that might signal the boundary between iris and pupil. The use of an infrared vidicon minimizes the effects of iris coloration on the contrast of the iris-pupil boundary. A single control is provided for the adjustment of the sensitivity of the contrast detection circuitry. Sensitivity is individually adjusted for each subject but, once adjusted, remains stable over long periods.

The second stage of imaging processing is the search for a semicircle of contrast points which together define the leading edge of the pupil. The diameter of this semicircle provides a reliable estimate of pupillary diameter. This measure is recomputed 30 times each second and is available for computer input in either analog or digital form.

The performance of the image processor may be evaluated by means of a video display of the processed image. Contrast points are indicated on the monitor as brightness-intensified sparkles. The extracted image of the pupil is visually indicated by a darkening of all raster lines passing through the detected pupil. Thus, if the pupillometer is functioning properly, the monitor displays a video image of an eye, with intensified points along the left iris-pupil boundary with a dark band tangent to the upper and lower boundaries of the pupil. Measurement quality can be assured by visual monitoring of the processor's display.

Recording Constraints: Accurate recording is essential to pupillometric measurement of workload since the pupillary dilations reflecting changes in central activation, although highly reliable and observable on single trials, are nonetheless exceedingly small. For this reason, other factors which affect pupillary diameter must be carefully controlled.

Chief among the non-cognitive determinants of pupillary diameter is the well-known light reflex, which reduces pupillary diameter as integrated retinal illumination is increased. The light reflex is very sensitive and the maximum amplitude of the response is several millimeters. For this reason the luminance of the visual field must be constant during measurement. In out experiments on visual information processing, we employed a computer-controlled CRT display in which task-relevant stimuli were presented for short (100-200 msec) periods. At all other times equiluminance random dot fields were displayed. Such control of the light reflex may not be possible if the subject is required to scan a complex visual field of varying luminance.

The momentary state of the occulomotor reflexes mediating convergence and accommodation also must be controlled as vergence movements and accommodation reflexively affect pupillary diameter. In our work with visual displays, the critical visual stimulus was placed several meters from the subject to relax accommodation and minimize convergence. At one time we were troubled with significant constrictions occurring in some subjects while viewing prolonged visual displays. We attributed these artifacts to uncontrolled vergence/accommodative movements and altered our task to utilize the more artifact-free brief presentations. Nonetheless, visual stimuli can be employed in pupillometric research, but a great deal of care must be taken in dealing with such materials.

These problems do not exist when auditory displays are employed. For this reason, presentation of information in the auditory mode is recommended whenever feasible.

Recording Artifacts: The video-scan pupillometer is one of the most accurate reliable, and trouble-free phychophysiological recording devices ever developed. Nonetheless artifacts in the pupillometric record do occur and must be dealt with before the data are analyzed.

The major sources of artifact are blinks and partial lid closures. In these cases, movements of the eyelid obscure a portion of the pupil, resulting in erroneous measurement. Such artifacts are easily observed in the pupillary record and are sufficiently obvious to permit automatic computer artifact detection if desired. In our own work, the raw pupillary data from a entire experimental session is stored on disk memory for later visual examination. Small artifacts are corrected by linear interpolation and data segments with large artifacts are discarded. This editing procedure is rapid and assures accurate pupillometric data.

Another major source of artifact lies in the contrast detection threshold established for certain subjects. If the illuminator is improperly focused, or if the subject has long drooping eyelashes, the recognition of the upper pupil boundary may be uncertain. This results in a characteristic jitter in the pupillary record. When this occurs the source of the difficulty should be corrected. Data segments containing such jitter should be discarded.

PUPILLARY CHANGES IN HUMAN INFORMATION PROCESSING

There is a large body of experimental evidence that suggests that pupillary dilations under controlled conditions reflect with high accuracy the momentary level of load placed upon the human nervous system by information processing tasks of varying difficulty, content and complexity. These data have been reviewed recently by Goldwarer (1972) and Janisse (1974). In this section I shall present a series of experiments, primarily from my own laboratory, that illustrate the sensitivity of pupillometric measures to momentary changes in processing load which are in some degree relevant to the applied problem of pilot workload evaluation.

Memory Load: One component of pilot workload is the demand placed upon short-term memory in verbal communication with other aircraft or ground sites. Detailed verbal instructions for example, must be accurately retained. The limitations of short-term memory are well known to psychologists and human factors engineers alike. Pupillometric measures provide a means of quantitatively assessing the physiological load placed upon an individual by verbal informantion of varying amounts and complexity which is to be retained for short periods of time.

Kahneman and Beatty (1966) presented the first pupillometric analysis of the processing demands encountered in a short-term memory task. Figure 1 presents pupillometric records obtained during a short-term memory task in which strings of 3 to 7 digits were auditorily presented at the rate of 1 per sec. Two seconds after the last digit was heart, subjects were required to repeat the digit string at the same rate. It is apparent from Figure 1 that the momentary degree of pupillary dilation accurately reflects the cognitive workload imposed by the short-term memory task. Pupillary diameter increases in a linear fashion with the presentation of each digit, reaching the maximum in the 2-sec pause preceding report. As digits are unloaded from memory during report, pupillary diameter decreases with each digit reported, reaching baseline levels after report of the final digit. In unpublished work, it was determined that if the subject were requested to repeat the string a second time immediately after reporting the final digit, the pupil immediately dilates to the peak diameter for that string and then decreases with each digit spoken until the entire string has been reported for the second time. The magnitude of the pupillary dilation at the pause between input and output in Figure 1 is an increasing function of string length. Beatty and Kahneman (1966) demonstrated that a similar pupillary function is obtained when a string of items is recalled from long-term memory for report: On request to report, a large pupillary dilation is observed as information is retrieved from long-term memory (see Figure 2). As each digit in the string is reported, pupillary diameter decreases, reaching baseline levels at report of the last digit. Thus it appears that the limited capacity portion of the human information-processing system may be loaded from either long term memory or environmental stimuli and that the pupillometrically measured workload is similar in both of these cases.

Memory load is also determined by the difficulty of the to-be-remembered information. Remembering unrelated nouns requires more capacity than remembering a string of single digits of equal length, as measured by the difference in memory span for the two types of items. Figure 3 shows the pupillometric data obtained for strings of four items of different types. The smallest dilations are observed for strings of four digits that were to be simply repeated. Larger dilations were apparent for the string of four words, indicating that both item difficulty and number of items determine workload in the memory task. The largest dilations were obtained for the subjectively most difficult task of transforming each of the four digits by adding one before report. These data provide strong support for the idea that task-induced pupillary dilations provide a physiological index of the momentary level of workload imposed by a memory task.

This idea was subsequently confirmed in an experiment by Kahneman, Beatty, and Pollock (1967) in which both pupillometric and behavioral interference methods were utilized to assess workload in the four-digit add-one memory transformation task. Using a secondary task of visual target detection, it was found that the behavioral estimate of workload and the pupillometric measure of physiological load were in exact agreement. A series of controls ruled out any peripheral interference of the pupillary dilations themselves on performance of the secondary task. In comparing the two data sets, the pupillometric data was for more detailed than the behavioral data, required fewer trials to obtain, and was of considerably lower variance.

Decision Processes: Even simple decision processes appear to impose some workload on the cognitive system as indicated by pupillometric measures of activation. For example, Simpson and Hale (1969) measured pupillary diameter in two groups of subjects who were required to move a level to one of four positions. In the decision group, subjects were told at the beginning of each trial that either of two directions was permissible (e.g., front or left). Seven seconds later a response cue was presented and the subject initiated one of the two movements. In the no-decision control group, subjects were instructed exactly as to the desired movement on each trial (e.g., front). Pupillary dilation in the post-instruction pre-response period was larger and more prolonged for thos subjects who had to choose between two movements before responding.

Substantially larger pupillary dilations are observed to accompany more difficult decision processes. In an experiment reported by Kahneman and Beatty (1967), listeners were required to determine whether a comparison tone was of higher or lower pitch than the standard. Clear pupillary dilation occurred in the 4-second decision period between presentation of the comparison tone and the response cue. The amplitude of this dilation varied as a direct function of decision difficulty, the difference in frequency between the standard (850 Hz) and comparison tones. This relation is shown in Figure 4, which presents both the amplitude of dilation in the decision period and the percent decision errors as a function of the frequency of the comparison tone. These dilations were highly reliable and did not habituate over the experimental session. Pupillary dilations during decision appear to vary as a function of cognitive workload, as inferred from task parameters and performance data.

Complex Reasoning: More complex cognitive functions not unexpectedly impose a major load upon the human nervous system during their execution. This may be most easily observed in the laboratory using mental arithmetic tasks. Such tasks may be regarded as directly analogous to other types of complex reasoning tasks that may occur in man/machine interactions.

Pupillary dilations accompanying complex problem solving appear to be related directly to the difficulty of such processing, although behavioral assessments of workload have not yet appeared for these types of cognitive tasks. For example, Hess and Polt (1964) examined pupillary movements as multiplication problems were solved mentally. Pupillary diameter increased during the period preceding solution, and related to presumed problem difficulty. Payne, Parry, and Harasymiw (1968) also report a monotonic relation between mean pupillary diameter and problem difficulty, but note that this relationship is markedly nonlinear with respect to difficulty scales based upon percent correct solution, time to solution or subjective rating of difficulty. Pupillary diameter in mental multiplication appears to peak rapidly as a function of difficulty, with more difficult problems requiring more time until solution is reached. This suggests that cognitive capacity is quite fully taxed in complex mental arithmetic problems so that the workload per unit time remains relatively constant as problem difficulty is increased over moderate levels, but that the total time to solution is increased.

These investigations using the older photographic methods of pupillometric measurement were not able to discern the fine temporal structure of complex reasoning tasks which is clearly evident when more detailed video-scan pupillometry is employed. Ahern and Beatty (in preparation), as part of a study of individual differences and cognitive load, presented subjects with multiplication problems at three levels of difficulty. The problems were computer-controlled using acoustically-presented digitized speech stimuli. These data are summarized in Figure 5. Clear dilations may be observed in all cases at the presentation of the multiplicant (a single digit, a low two digit number or a high three digit number). This dilation quickly subsides and the pupil returns towards basal levels until the multiplier is presented, at which point a major dilation is observed. The duration of this dilation is related to problem difficulty being more prolonged for more divvicult problems. These data suggest that pupillometric methods not only may serve to measure the workload associated with a single task or function, but also to measure the time course of that load with some degree of precision.

Other types of complex problem solving tasks show similar relationships between pupillary dilation and problem difficulty. For example, Bradshaw (1968) has reported that larger pupillary dilations accompany the solving of more difficult anagrams, and that these dilations are maintained until solution is reached.

Summary: Pupillometric measurements have now been obtained in a variety of simple information-processing tasks under laboratory conditions. They appear uniquely sensitive to subtle differences in processing load obtained in these tasks. Processing load appears to increase the activation of brainstem arousal systems in measured amounts. These activation responses are of short duration, of an extent that accurately reflects load, and occur at short latency. The responses do not habituate, and therfore may be assumed to reflect a fundamental physiological response to increase in cognitive workload. As such, they suggest an alternative to traditional methods of quantifying workload, a possibility that is explored in the following section of this report.

FUTURE APPLICATIONS OF PHYSIOLOGICAL MEASURES TO THE ASSESSMENT OF PILOT WORKLOAD

No investigation has yet been published in which pupillometric methods have been employed in the measurement of pilot workload. Perhaps the most direct application of these methods to practical performance assessment is Peavler's (1974) use of pupillometric measures to assess fatigue in telephone operators after working full shifts on different types of computer-based information retrieval systems. Peavler found that the more automated method, which was both more efficient and more taxing, resulted in greater operator fatigue, as indexed by mean decrease in pupillary diameter from pretask to posttask measurements. Thus, Peavler was not concerned with the question of task-induced pupillary dilations and instantaneous workload levels, the topic of the present report.

The body of research summarized above certainly makes a theoretical contribution to the study of workload, suggesting that workload can be measured by a physiological response to task load, rather than by behavioral interference or subjective report. In my opinion, these methods may be of practical consesequence as well.

The most natural application to the problem of pilot workload would seem to be in the area of design of equipment and pilot procedures, in which the workload parameters of each of several design options might be assessed separately using experimental methods similar to those outlined above. Here, one might ask questions concerning optimal information formatting to determine a communication structure that minimizes operator load. The method is particularly well suited for the design of the more cognitive components of the pilot's task, analogous to the mental arithmetic experiments described above. It is precisely this aspect that would seem to be most difficult to measure by conventional workload assessment procedures.

One could conceivably construct a simulator in which pupillometric measurements might be made to test workload in a more realistic environment. However, in my opinion, the problems of adequate control of visual input in such a situation would seriously impede its usefulness. As mentioned above, strict control of visual input is necessary for the pupillometric measurement of workload as the large magnitude changes in pupillary diameter that are produced during a visual scan of a non-homogeneous visual field introduce serious artifacts in the pupillometric record. Until such problems are solved, the use of pupillometry in more natural environments will be restricted at best.

Finally, some attention should be paid to the use of other physiological measures such as the EEG in the assessment of workload effects. An inspection of the current literature is not promising in this regard, as no large magnitude and robust relations between EEG and workload have been reported despite a reasonable amount of experimental work devoted to this problem. The development of an EEG measure of workload would be of some practical interest, as the EEG is not dependent on small changes in visual input as is the pupil. The question of an EEG measure of workload is presently being pursued in my laboratory under ONR support. We are using the mental arithmetic and short-term memory tasks which have such strong and reliable effects on autonomic indicators of load, including the pupil. Pupillometric data are also being analyzed. EEG data is being systematically recorded from each of the 19 sites in the Ten-Twenty recording system (Jasper, 1958) and stored for subsequent analysis. By proceeding in a systematic manner in the analysis of the EEG and by continuing use of the pupillometric measures to assess the effectiveness of the manipulations of processing load, we hope to finally discern the central signs of processing load which are so clearly observable in the autonomic periphery.

In summary, physiological methods provide a unique alternative to the traditional, but in various ways unsatisfactory, methods of workload measurement, subjective estimation and behavioral interference with a secondary task. Of the physiological measures, the task-evoked pupillary responses provide the clearest indication of both the degree of load imposed by a particular task or function and the fluctuations of that load over time. Although some restrictions are necessary to insure accurate pupillometric recordings, the use of pupillometric methods for workload assessment would seem to be feasible, particularly in evaluating the load imposed by complex cognitive tasks.

REFERENCES

Ahern, S. K. and Beatty, J. Activation and intelligence: Pupillometric correlates of individual differences in cognitive abilites. In preparation.

Goldwater, B. C. Psychological significance of pupillary movements, Psychological Bulletin, 1972, 77, pp. 340-355.

Hess, E. H. and Polt, J. H. Pupil size in relation to mental activity during simple problem solving, Science, 1964, 143, pp. 1190-1192.

Janisse, M. P. (Ed.) Pupillary dynamics and behavior, New York: Plenum, 1974.

Jasper, H. H. The ten-twenty electrode system of the International Federation, Electroencephalography and Clinical Neurophysiology, 1958, 10, pp. 371-375.

Kahneman, D. and Beatty, J. Pupil diameter and load on memory, Science, 1966, 154, pp. 1583-1585.

Kahneman, D. and Beatty, J. Pupillary response in a pitch-discrimination task, Perception & Psychophysics, 1967, 2, pp. 101-105.

Kahneman, D., Beatty, J., and Pollack, I. Perceptual deficit during a mental task, Science, 1967, 157, pp. 218-219.

Kahneman, D., Tursky, B., Shapiro, D. and Crider, A. Pupillary, heart rate and skin resistance change during a mental task, Journal of Experimental Psychology, 1969, 79, pp. 164-167.

Moruzzi, G. The sleep-waking cycle, Reviews of Physiology: Biochemistry and Experimental Pharmacology, New York: Springer-Verlag, 1972.

Payne, D. t., Parry, M. E. and Harasymiw, S. J. Percentage pupillary dilation as a measure of item difficulty, Perception & Psychophysics, 1968, 4, pp. 139-143.

Peavler, W. S. Individual differences in pupil size and performance, In M. Janisse (ed.), Pupillary dynamics and Behavior, New York: Plenum, 1974.

Simpson, H. M. and Hale, S. J. Pupillary changes during a decision-making task, Perceptual and Motor Skills, 1969, 29, pp. 495-498.

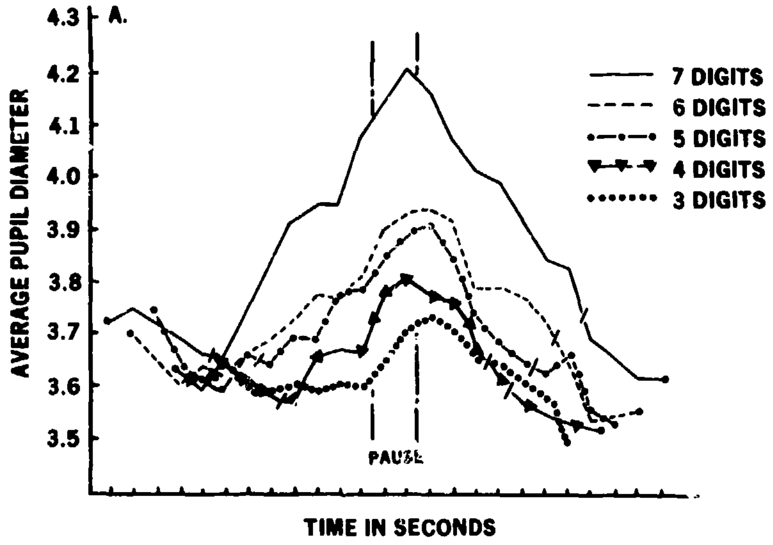


Figure 1. Average pupillary diameter during presentation and recall of strings of 3 to 7 digits, superimposed about the two second pause between presentation and recall. Slashes indicate the beginning and the end of the memory task.

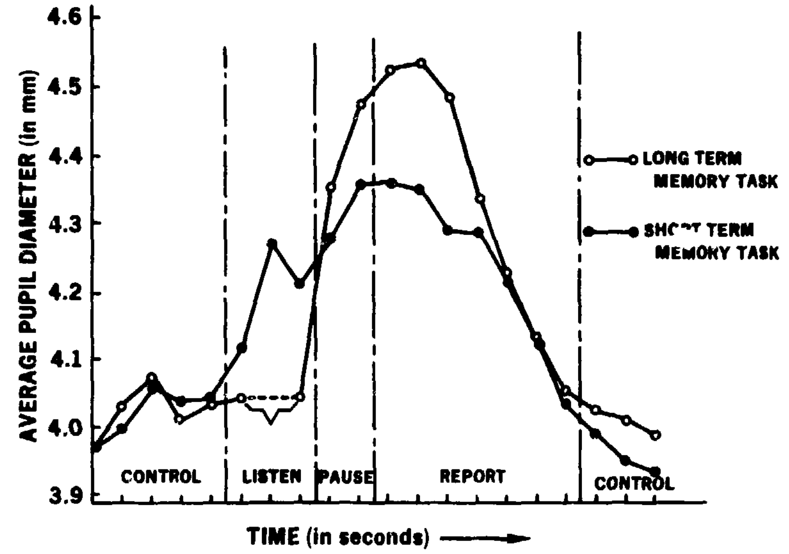


Figure 2. Average pupil diameter for five subjects during presentation and report of seven-digit telephone numbers from short-term and long-term memory. The long-term memory function is broken above the brace, with both points representing the same pupillary measurements.

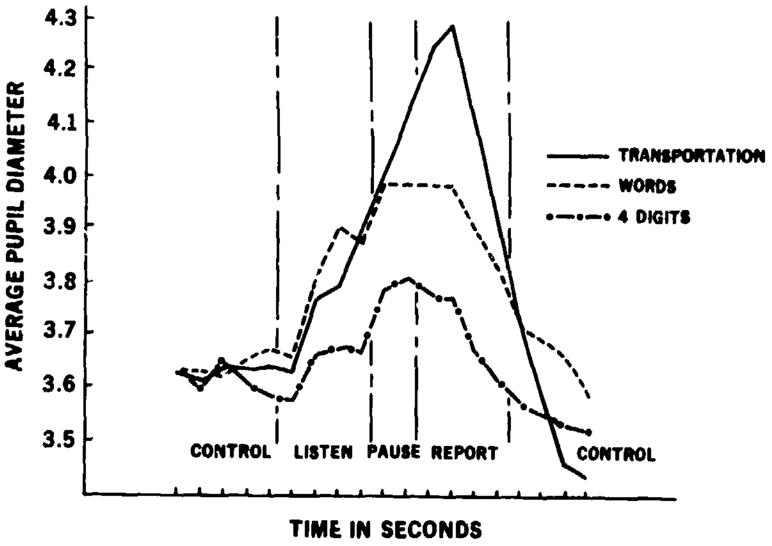


Figure 3. Pupillary diameter during presentation and recall of four digits, words and a digit transformation task.

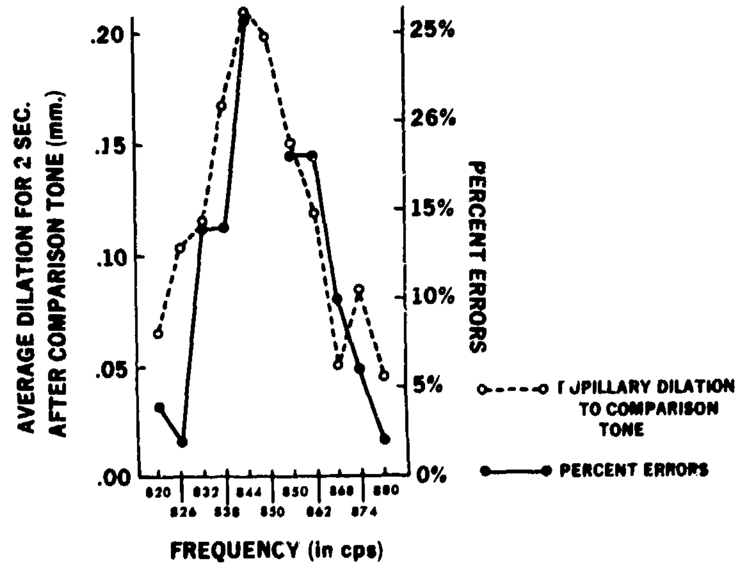


Figure 4. Average pupillary dilation during the decision period and percent errors as a function of the frequency of the comparison tone. The frequency of the standard was 850 cps.

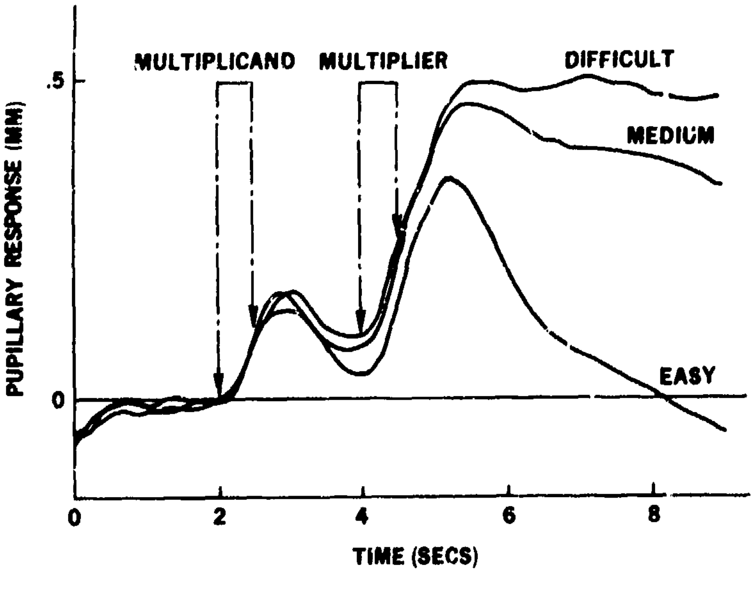


Figure 5. Averaged evoked pupillary responses in a mental multiplication task of three levels of problem difficulty.

AIRCREW PERFORMANCE RESEARCH OPPORTUNITIES USING THE AIR COMBAT MANEUVERING RANGE (ACMR)

by

Clyde A. Brictson, Ph.D.
Anthony P. Ciavarelli, M.A.
Dunlap and Associates, Inc.
Western Division
7765 Girard Ave., Suite 04
La Jolla, California 92037

ABSTRACT

Three years of aircrew performance measurement related to air combat effectiveness using the Navy's Air Combat Maneuvering Range (ACMR) are presented as evidence of ACMR's research potential. Performance assessment methods used to evaluate pilot proficiency are described. The aircrew assessment methods have been used to identify squadron performance differences, evaluate competitive exercises, and provide diagnostic training feedback to operational users. The use of continuously recorded quantitative measures from systems such as ACMR should stimulate more aircrew performance field research ideas. The avilability of objective performance criteria promises to be of substantial benefit to both the operational user and the research community in such areas as pilot selection and training, fleet combat readiness, and pilot workload and stress.

INTRODUCTION

Background: The selection, training and assessment of military aviators, and problems associated with the acquisition and retention of flying skill, have occupied aviation psychologists for over 30 years (Thorndike, 1974). The major problem in this line of research has been, and continues to be, the lack of objective criteria (North and Griffin, 1977) for evaluating the effectiveness of aviation training in general, and aircrew proficiency in particular. Traditionally, the use of subjective estimates has provided the only means to assess training progress in acquiring and maintaining aviation skills.

The recent growth of the Navy's Air Combat Maneuvering Range (ACMR) has provided a unique opportunity to obtain objective measures of aircrew performance that have not been avilable in the past. For the past three years the authors have been involved in a research program to develop objective aircrew performance criteria from ACMR quantitative output measures. Two technical reports (Brictson and Ciavarelli; 1976 and 1978) have been written which detail the technical approach, performance assessment methods, and preliminary results of aircrew performance measurement on selected training objectives. The ACMR criterion development research is sponsored by the Navy Aerospace Medical Research Laboratory, Pensacola, Florida in order to provide them with performance criteria to validate, among other things, vision laboratory results, aircrew selection practices, and training effectiveness. The availability of such criteria, however, has perhaps more far reaching implications for the expansion of aircrew research efforts in an operational environment.

Air Combat Maneuvering Range (ACMR): The ACMR is a sophisticated training facility acquired by the Navy and now in use to train fighter aircrews in air-to-air combat. The system is designed to train aircrews in actual combat maneuvers and in recognition of weapon delivery boundaries. ACMR provides data display features which greatly enhance air combat debriefs, and provide a rich source of continuously recorded quantitative measures. Some of the capabilities of ACMR include the following:

1. Real-time tracking of aircraft engaged in air combat training in a specified airspace,
2. Video tape playback of flight history data, complete with pictorial display of the air-to-air engagement and voice transmissions,
3. Both digital and graphic hard-copy printouts of flight instrument data, interaircraft positions, cockpit view of engaged aircraft, mission data, and
4. Computer generated estimates of weapon launch outcomes.

ACMR as a system enables training and research personnel to monitor in real-time various air combat training exercises, and through exercise replay, provides the opportunity to review, debrief and evaluate pilot tactics, decisions, and weapon delivery accuracy. In addition, expected ACMR advances are designed to obtain measures in attach mission roles as well. Planned system augmentation will cover no-bomb-drop scoring, mine laying operations, anti-radiation and electronic warfare missions The whole array of operational missions and their slow-motion replay will soon be within the province of aviation research teams to better understand and resolve the complexities of pilot/aircraft matchups.

RESEARCH REVIEW

In-Flight Assessment Methods: Our technical approach (Brictson, Ciavarelli, et. al., 1977) describes an appropriate systems framework, training content, and performance assessment methodology for the development of reliable and valid ACMR criterion performance measures.

Measures from over 600 ACMR dog fights have been obtained across a variety of aircraft and weapon systems, and under varying training missions and operating conditions. A performance assessment methodology was developed and used to evaluate aircrew and squadron air combat performance. The performance assessment methods include analysis of engagement outcomes (wins, losses, draws) as well as task accuracy measures associated with successful weapon delivery. Recently, we have developed metrics from the analysis of antecedent events (i.e. radar contact, initial visual acquisition, first engagement short) in order to estimate the probability of any given outcome, given certain antecedent conditions.

Collectively, these assessment methods provide a complete measurement system for estimating aircrew and unit proficiency in all aspects of air combat maneuvering. We will soon be able to provide longitudinal and objective data on all critical phases of air combat maneuvering.

Performance Results: Since ACMR instrumentation provides so many output measures there is a range of indiscriminant selection of candidate measures of performance. We ran across many occasions where it was tempting to measure 'everything that moves,' but we chose instead to look at the statistical and practical aspects of the data--recognizing fully that if your results do not make sense to the operational community they will not be used.

To arrive at a reduced set of candidate measures we first identified thirteen air combat training objectives and, using various logical and documentary criteria, selected weapon envelope recognition as the most critical to success. A comprehensive statistical analysis, using ANOVA, multiple correlation and discriminant analysis, resulted in the selection of two statistically and practically significant variables from the multitude of measures available on ACMR. In the final analysis a single error score, which was defined as a deviation from ideal weapon delivery boundary zones, proved to be the most promising measure of envelope recognition task accuracy. Based on that conclusion we have now developed empirical distributions of these error scores for high and low pilot performance and experience continous for use as baseline data to evaluate any future training innovations or system improvements in envelope recognition.

In general, the progress of ACMR performance criteria development has produced some very promising results. We have, for example:

- Identified key variables related to successful weapon delivery,
- Developed preliminary criteria for evaluating aircrew performance in envelope redognition, and
- Devised scoring metrics based on engagement outcomes and task accuracy measures which have demonstrated their effectiveness in discriminating known performance differences.

More importantly, we now have in-hand a list of statistically and practically significant variable which not only account for the major portions of variance related to air combat success but are also -- and this is critical to measurement success -- understood and accepted by the operational user, i.e. pilots and training officers.

Efforts are continuing to further refine and expand performance assessment techniques, and to establish the statistical integrity of the data base for ultimate application in support of both operational training and for validation of ongoing aviator research. While the training application of these data are readily acknowledged, the research aspects and potential have yet to be realized in the research community at large. We hope that this brief foray will entice other aviation research teams to utilize the tremendous capability now available in ACMR systems emerging around the world.

A NEW ERRA

For the past 30 years aviation psychologists, given the lack of objective operational measures, have been forced to do research designed primarily to enhance the reliability and validity of subjective and second order 'criterion measures.' Usually the criterion measures so developed rested on the use of flight instructor subjective estimates or peer training which met with various degrees of success. With the arrival of training systems such as ACMR aviation psychology has crossed the treshold into a new era.

The avilability of continuously recorded and objective output measures, along with on-line computer analysis and display, present the researcher with a completely new capability to evaluate performance 'on the job.' Although much remains to be done to demonstrate the generalizability of initial performance assessment methods developed to date, the methods have already been successfully demonstrated across small samples and show remarkable promise.

The utility of reliable and valid objective performance criteria can not, and should not, be underestimated. From an operational view point, the measures are essential for judging the progress of ACMR training, estimating aircrew proficiency levels, and for determining the combat readiness of operational units.

On the other hand, the research community now has at its disposal operational measures as potential validation criteria for ongoing aviator selection, training and research programs. The air combat mission is most certainly one of the most demanding tasks in terms of skills required and stresses experienced. ACMR provides a vehicle for the field validation of research directed at understanding the acquisition of these skills and the conditions under which they may be enhanced or degraded.

Going hand-in-hand with operational measures related to aviator combat missions is the present availability of aircraft carrier final approach landing scores (LPS) which have already been tested and validated in the fleet (Brictson, et. al., 1973). ACMR and carrier landing measures, used independently or in combination, provide a unique opportunity to support ongoing research related to the selection, training, and performance effectiveness of Navy aviators.

Given the avilability of operational performance measures, researchers can more effectively address some of the questions that have arisen over the history of aviation research. Some of these questions are of very high priority to the nation's defense in general, and to Naval aviation in particular. For example:

1. When is an aircrew, squadron or fleet considered combat ready in ACM?
2. What are the effects of ACM on pilot physiological responses?

3. How frequent is ACM practice required to maintain proficiency?
4. What is the range of pilot stress tolerance to ACM missions?
5. What is an acceptable standard of ACM performance for different training levels?
6. How can operational performance measures be used to select top aviators?
7. And finally, what are the effects of sustained operations, prolonged duty hours, and operational workload on the performance effectiveness of Naval aviators?

The answers to these and other operationally relevant questions can now be obtained given access to on-line performance measurement systems such as ACMR and represents an unequalled opportunity for aircrew performance research.

RESEARCH OPPORTUNITIES

Of prime importance to research workers dealing with aviator workload, stress and fatigue is the intriguing notion of an on-line pilot monitor system during air combat missions. Long considered to be one of the more stressful and demanding pilot tasks, an air-to-air engagement taxes the pilot physically, mentally and perceptually. The possibility of complimenting on-line pilot performance measures with on-line physiological measures such as heart rate, blood pressure, etc. would provide an ideal arrangement for the research team interested in validating laboratory notiont of stress, fatigue or workload in an operational 'real world' environment.

A word of caution is advised. Some research teams used to the controls and precision design of experiments in the laboratory will be limited in their attempts to control the real world. But that is exactly the point. Many laboratory studies stress the statistical significance of results without strong support for practical or operational significance. In pilot workload, for example, the amount or severity of workload in either a 24-hour or flight segment is certainly useful to 'describe' the environment but does not by itself have any practical significance unless it can be related to performance effectiveness, short or long term. Our physiological reactions to stress or workload can assuredly be measured but it is only in the context of their relation to performance that they acquire operational significance.

With the advent of sophisticated instrumentation systems like ACMR and the concurrent development of performance criterion measures the final building block in field calibrated research is in place. All that now remains is the historical challenge of innovative and understandable test designs that can answer operationally significant problems.

Out own approach in ACMR is to provide, first of all, valid and reliable performance criteria. Secondly, we want to obtain a longitudinal performance data bank based on pilot biographic, experience, biochemical, sleep, mood and workload components. Third, and most important, is our interest in having a field laboratory that can provide an arena to explore, define and predict the influence of pilot temporal variables on aviation performance effectiveness.

The ACMR system, while now prevalent in the contental U.S.A., is also being made available to NATO nations for training purposes at a location in Sardinia. NATO scientists, ideally, could have access to the performance data through part-time use of the facility for research purposes. Many of the papers recently discussed at the 1977 Cologn AGARD Panel meeting on pilot workload could benefit from on-line performance measurement data such as that provided by ACMR. In addition to land based ACMR systems there is a strong likelihood that ACMR, with its vast potential for tapping continuously many aspects of pilot performance and physiological responses, will also be available at sea, aboard various U.S. Navy aircraft carriers. If that planned installation occurs then the use of ACMR for research purposes could greatly expand due to greater availability of ACMR facilities at sea and ashore. Regardless, it is now possible to obtain from ACMR reliable and valid operational meausres of air combat maneuvering. Such measures should provide a wealth of opportunity for research teams from NATO, USN and USAF communities.

REFERENCES

1. Brictson, C. A., Burger, W. J. and Wulfect, J. W. Validation and application of a carrier landing performance score: The LPS. Inglewood, California: Dunlap and Associates, Inc., March 1973.

2. Brictson, C. A., Ciavarelli, A. P. and Jones, T. N. Development of aircrew performance measures for the air combat maneuvering range (ACMR) (U). Pensacola, Florida: Naval Aerospace Medical Research Laboratory Technical Report L53001, June 1977. (CONFIDENTIAL)

3. Brictson, C. A., Ciavarelli, A. P., Pettigrew, K. W. and Young, P. A. Performance assessment methods and criteria for the air combat maneuvering range (ACMR): Missile envelope recognition (U). Prepared for Naval Aerospace Medical Research Laboratory, Pensacola, Florida. Prepared under contract N61339-76-C-0082, November 1977, (In press.) (CONFIDENTIAL)

4. North, R. A. and Griffin, G. R. Aviator selection 1919 - 977. Pensacola, Florida: Naval Aerospace Medical Research Laboratory, Special Report 77-2, October 1977.

5. Thorndike, R. C. Research problems and techniques. Washington, D.C.: U.S Government Printing Office, Army Air Forces Aviation Psychology Program Research Reports, Report No. 3, 1947.

Note: The research reported in this paper was completed under Navy contract N61339-77-C-0167. The opinions expressed here are those of the authors and do not represent official Department of the Navy policy.

SPEECH PATTERNS AND AIRCREW WORKLOAD

R Cannings
Royal Air Force Institute of Aviation Medicine
Farnborough, Hampshire, United Kingdom

SUMMARY

The use of speech patterns in the analysis of workload is examined. The rather sparse amount of research effort expended in this field is reviewed in terms of a simple model of speech production and the applications of current analysis techniques are considered.

INTRODUCTION

There is much intuitive evidence to suggest that high workload or stress may change the fundamental characteristics of speech, and so although the voice may not exhibit obvious variations during normal flight profiles, a search for change in speech may prove to be a worthwhile approach in the investigation of workload in air operations. However, central to the possible use of speech patterns is the requirement to reduce complex speech data to parameter sets of a manageable size, and to relate these sets to the psychological and physiological state of the pilot. Optimum choice of parameter sets constitutes a difficult task, but there is an ever increasing literature concerned with speech processing which provides many techniques of analysis.

Reliable voice parameters may be extracted from the relatively poor quality speech of existing flight communication channels, and so speech patterns may prove to be useful, as they overcome the need for subject instrumentation and data collection (see for example Refs 1-3). Correct choice of speech parameters may make it possible to assess changing workload patterns, and this may be important in the military environment where rapid fluctuations in workload and stress are encountered, and where many other methods, such as those which rely on biochemical analysis (see for example Ref 4), may be of little value.

THE SPEECH WAVEFORM

There are several questions raised by the use of voice analysis in the investigation of workload in aircrew. In common with many other techniques, the processes underlying the variations in voice parameters are uncertain. An effect may be produced in response to endocrine changes when it is likely that the response time will be long in relation to the duration of the workload and the induced stress. If on the other hand, the change arises essentially from increased neurophysiological activity, then the response time of the effect will be rapid. In each case, the basic mechanisms of speech production would be similar and are documented in the literature (see for example Ref 5), but there is little information concerning the processes which may invoke variation under high workload.

Further, vocalization is a conscious process. The majority of fundamental parameters which characterise a given speaker may be split into those which transmit the information content of the spoken word, and those which do not contain this semantic information. The former group, for example formant frequencies, vary according to the particular utterance, but over long periods of time, as changes due to semantic information average out, the range of variation is beyond direct conscious control. The latter group contain measures such as fundamental frequency or pitch, which are responsible for intonation. Western languages do not require changes in pitch to transmit semantic information, but conscious control may produce significant short term variations. Although the nature of the short term changes may be of interest, mean fundamental frequency in the long term is again thought to be beyond direct conscious control.

In view of these considerations it is worth reviewing the way in which a set of parameter estimates should be used. As an illustration, a voice parameter from a pilot is measured during the course of a single flight, and it is assumed that, initially, there is no knowledge of the influence of high workload. The time course distribution of the estimates of this parameter during the flight will depend upon the times at which the pilot chooses to speak. The estimates are likely to be corrupted by noise due to poor recordings, problems of measurement and random or conscious variations in the pilot's voice. The absolute values are likely to be of little value, but relative changes through the flight profile may be of greater interest. Statistical methods are available to establish whether any trends exist, and to test if a particular aspect of the flight profile shows a significant change. Similar methods would be applicable if estimates of the same speech parameters from an unstressed situation were available, either in flight or on the ground. Measures of the relative change could be quite useful, given a sufficient knowledge of the flight profile which would identify times of high workload. However, utterances are often short and randomly dispersed through the flight profile, and so in preliminary studies, it would be desirable to correlate with other physiological data. Such data are easier to gather from transport flights because of the ease of instrumentating the pilot.

Data from many flights are necessary to establish the existence or otherwise of specific trends related to high workload, and if trends are found, it would be necessary to establish whether they are reproducible in different pilots. Studies of changes in the voice under stress have demonstrated wide inter-subject variability (Refs 6-7), and these observations raise a much broader question, concerning the way in which various parameters from different pilots could be evaluated for indications of high workload. Given speech data from a single pilot, it is possible to use techniques of ever increasing complexity until changes are found which significantly reflect high workload situations. Such studies are time consuming, and, even so, the final technique may or may not be relevant to other pilots. Alternatively, more simple techniques may be applied to data from several pilots in an effort to establish trends across pilots. Intuitively, the latter approach is felt to be more realistic, even if some aspects of speech requiring involved methods of extraction are ignored.

It is also necessary to bear in mind the techniques of analysis which are available, and how they relate to the quality of the available data. For example, inverse filtering methods may be used to obtain the shape of the glottal pulse, which is the basic element in the quasi-periodic waveform responsible for fundamental frequency. To do this, speech recordings of adequate quality are essential (Ref 8). Finally, when considering the existing literature, the relevant information is sparse, but section 4 offers a brief review, in which it is hoped the problem and various approaches may be put into perspective.

MODEL OF SPEECH PRODUCTION

Though a description of the physiology of speech production would be out of place in this review, nevertheless, it is worthwhile to describe the genesis of the speech waveform. This presents the opportunity to define aspects of the waveform which can be measured, and which may reflect a high workload situation. Figure 1 illustrates the way in which speech may be broken down into its phonetic components.

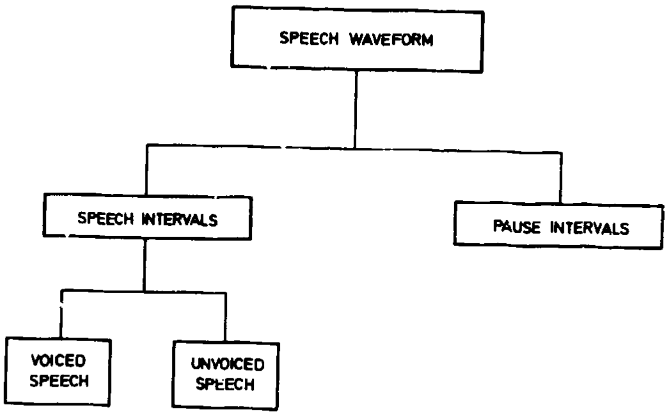


Fig 1
Phonetic composition of the speech waveform

Any utterance consists of periods of vocal activity and non-activity, known respectively as speech intervals and pause intervals. In isolation, the latter are of no interest, but together they provide information on the speech pause ratio, and on the overall rate at which the pilot is talking. This apparently trivial point is of some importance, especially when obtained as part of an analysis of the speech waveform envelope shape. The envelope shape reflects the duration of phonetic segments as well as overall articulation, or the precision with which different sounds are produced. Such measurements may contain information on high workload situations (Ref 6), even though it is likely that this information only reflects changes in the pattern of respiration. Unfortunately, the discontinuous nature of cockpit communication rarely provides a speech epoch of sufficient length for this form of analysis.

Figure 1 also shows that speech intervals may be divided into voiced and unvoiced segments. This broad classification is dependent upon the presence, or absence, of vocal chord activity. Speech intervals can be described by the model illustrated in Fig 2, which is based on the acoustic theory of speech production (Ref 5). In digital form, the model has found extensive application in computer based analyses which extract voice parameters from the speech waveform (see for example Ref 8-12). The first part of the

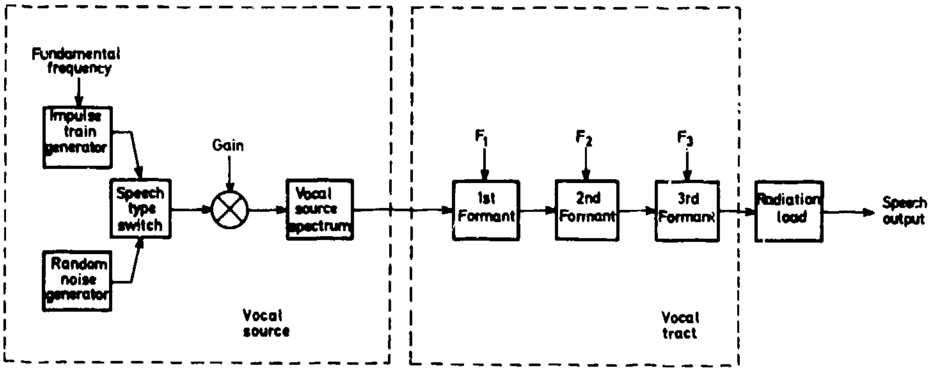


Fig 2
Model based on the acoustic theory of speech production

model comprises two possible excitation sources and a source filter. The type of speech depends largely on the excitation source, with the random noise generator producing unvoiced sounds or fricatives. In actual speech, a constriction is formed in the vocal tract and air is forced through it, generating turbulence, and hence noise. A combination of the random noise generator and impulse train generator can produce the so called voicedfricatives, and plosive sounds are created in a transitional phase between pause intervals and voiced or unvoiced speech intervals. However, none of these three types of sound has any simple application to the current problem.

More important are the vowel sounds, or voiced speech sounds, which are derived from the quasi-periodic impulse train generator. The instantaneous period of the pulses defines the fundamental frequency of the voice, which usually lies in the range 80-300 Hz. Many speech analysis-synthesis systems are based on voiced speech models, and as a consequence the vocal source spectrum and vocal tract resonators need only be considered in relation to this type of excitation. The concept of an impulse train is an idealization, because practically, puffs of air are released into the vocal tract by vibration of the vocal chords. The shape of each puff, known as the glottal pulse, is determined by the vocal source spectrum, and is largely dependent upon the state of the larynx and vocal chords. Figure 3 illustrates a simple electrical circuit which represents the sub-glottal system. The bronchi and trachea are represented as T sections, driven by a voltage representing the alveolar pressure, P_{alv}. Elastic recoil in the lungs, ie charge on the lung compliance capacitor C, is sufficient to produce normal expiratory airflow. The lung tissue resistance is negligible and may be ignored. During phonation, however, inspiratory muscle activity will produce a negative, (subatmospheric) intrapleural pressure and impede expiration. This produces a highly regulated expiratory flow through the glottal area. Usually airflow is small and so the sub-glottal pressure, P_s, and alveolar pressure are nearly the same. The resistance and inductance, R_g and L_g respectively, represent the variable area glottal orifice. For voiced sounds in the normal pitch range, the resistive term is dominant. In the context of stress analysis, the properties of this model are conveniently summarized by the fundamental frequency of vocal chord activity and vocal source spectrum, as viewed from the vocal tract.

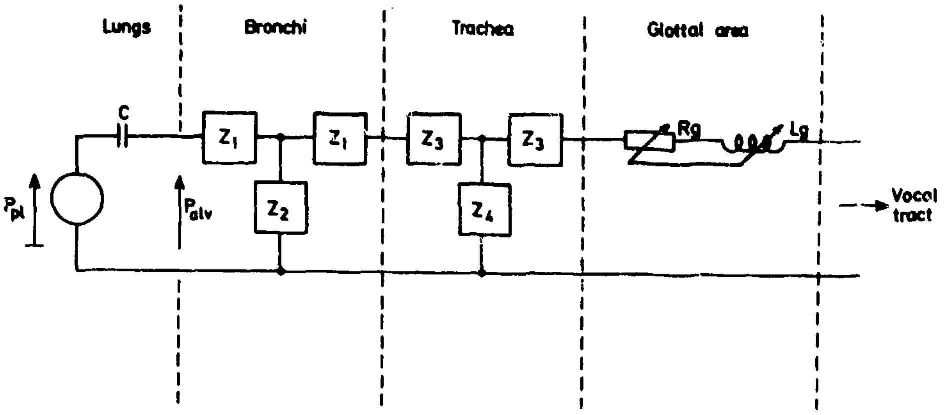


Fig 3
An electrical circuit representation of the sub-glottal system

The physical factors which control fundamental frequency and the vocal source spectrum are closely related, and it has been suggested that they are important in evaluating high workload situations (Refs 7 and 13). This implies that the larynx is subject to the normal neuromuscular manifestations of stressful situations (Ref 14). Once again, the respiratory pattern may be important since an increase in sub-glottal pressure can change the shape of the vocal source spectrum by effectively narrowing the glottal pulse. Much of the literature concerned with stress in the human voice has used fundamental frequency as the indicator, but the glottal waveform has found little application, presumably, due to the computational complexities involved in its measurement.

The final feature of the vocal source is the gain multiplier, which has the effect of controlling the overall loudness of the speech signal. Except under controlled recording conditions, it is difficult to make use of amplitude information, or equivalently, absolute values in power spectra. There is an added complication in that an increase in the loudness of the voice is generally accompanied by an increase in fundamental frequency. In the final stages of a let-down, approach and landing, a possible increase in the fundamental frequency of the pilot's voice may not be due to high workload, but rather, an increase in voice loudness related to increased engine noise.

The vocal tract is characterised by a series combination of quasi-time-invariant linear band-pass filters, which are often termed formants. Each filter is characterised by a resonant frequency and a bandwidth. Realistic estimates of the resonant frequencies or formant frequencies may be obtained from relatively straightforward processing schemes, but this is not true of the formant bandwidths. However, speech experiments have suggested that with constant bandwidth parameters, time varying combinations of the formant frequencies offer realistic syntheses of the vowel sounds in sections of voiced speech.

The essential structure of each formant filter is shown in Fig 4. The filter transfer function is given as

$$H(s) = \frac{1}{s^2LC + sRC + 1}$$

and it follows that the spectral peak occurs at

$$\omega_{max} = \sqrt{\left(\frac{1}{LC} - \frac{R^2}{4L^2}\right)}$$

L, C and R summarize the properties of air motion in a cylindrical tube. L is an acoustic inertance and remains essentially constant. C is a compliance term which depends on the cross-sectional area of the vocal tract, while R is a viscous drag term, dependent upon both the cross-sectional area and the circumference of the vocal tract. Essentially, R controls the formant bandwidth and C the formant frequency. Both of these parameters vary relatively slowly and so the formant system may be regarded as invariant in terms of short-time analysis. In this context the process is considered stationary during periods of 20 mS or less.

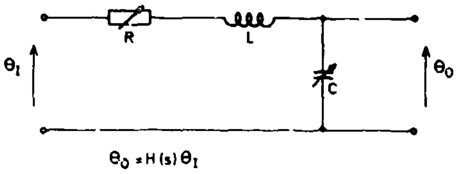


Fig 4
Electrical analogue of a single formant resonator

In theory there is an infinite number of formants, but in practice, three or four are sufficient to characterise a voice, although the acoustic theory of speech requires further filter elements for the correct representation of nasal consonants (Ref 5). In the male, empirical data suggest the first formant lies in the range 200-900 Hz, the second formant in the range 550-2700 Hz and the third formant in the range 1100-2950 Hz (Ref 10). Physically, the formant resonators comprise the cavities of the pharynx and oral and nasal cavities. The tongue, jaws and lips are also able to modify the low order formants. Some studies have considered possible interactions of stress with formants, essentially by examining changes in spectral balance within the formant frequency range. Such studies have been qualitative as well as quantitative (Ref 15-17), and will be considered in more detail later. Intuitively, however, since it is the glottal waveform which actually characterises the voice, the vocal tract will be of less interest, as it merely shapes the glottal waveform to produce semantic information (Ref 18).

The last component in the model of speech production is the radiation load. This filters the speech signal according to the way in which the vocal tract is coupled, via the mouth, to free space. In speech analysis applications it is often of greatest importance to obtain fundamental frequency and formant parameters, and so the characteristics of the vocal source spectrum and radiation load spectrum may be lumped together and removed from the speech signal, as both may be considered time invariant.

MEASURES OF STRESS IN THE SPEECH SIGNAL

In this section we are concerned with methods which have been used to establish whether stress modifies the speech signal.

Voice Micro-tremor. Although the previous section presented a model of speech production and highlighted the aspects of vocalisation which are likely to reflect stress, there is a further phenomenon known as voice micro-tremor, which does not fit into the scheme, but is nevertheless important. Tremor, or to be more specific, an 8-12 Hz modulation in the human voice, is a fairly recent discovery. Commercially, the phenomenon has found application as an extension to polygraph lie detector methods, and appears to have met with some success, at least in a well structured interview situation. One of the first devices offered simple strip chart recorder output, and required a skilled operator to interpret the results (Ref 19). A more recent device has a direct digital readout of stress level, but there is little technical information on its operation (Ref 20).

The proponents of such devices have attempted to explain the principles behind voice tremor, and, essentially, it is assumed that the muscles controlling the vocal chords exhibit the sort of tremor which accompanies activity in any of the voluntary muscles. It is postulated that this will cause slight rhythmic changes in vocal chord tension which will result in an inaudible 8-12 Hz modulation of fundamental frequency. Similarly, the muscles controlling the throat, lips and tongue are thought to be sensitive to the same kind of tremor, which will be reflected as a modulation within the first formant bandwidth. In a stress situation it is assumed that increased nervous activity causes muscle tension to increase throughout the body and, in the larynx at least, this will reduce the micro-tremor. In a high stress situation voice micro-tremor may disappear altogether. In view of the supposed mechanism of voice tremor, this is a rather curious observation, since other manifestations of muscle tremor appear to increase in the high workload situation (eg Ref 1). However, Inbar et al (Ref 21) have attempted to measure voice tremor and correlate it with muscle tremor in the area of the larynx. This technique was used to establish if voice tremor was due to mechanical "subresonances" in the vocal tract, or if it was generated by increased nervous activity. Their results suggest that micro-tremor is a frequency modulation of the glottal waveform, and that it is generated by nervous activity. Frequency modulations were also detected in the first and third formant bandwidths.

Although the commercial application of the voice tremor phenomenon is not in line with the current application, the underlying process would appear to make further investigation worthwhile. Several studies of commercial devices have been undertaken (eg Ref 14 & 22). Older and Jenny (Ref 14), carried out a comprehensive evaluation using the voices of astronauts from Skylab III and Skylab IV missions. Their conclusions suggested that the voice tremor principle, as exploited in commercial devices, would not detect any possible stress, at least in the Skylab situation. This may suggest that such devices are of real value only in the structured interview application. However, it should be pointed out that the commercial devices appear to be of simple design, and since we are not aware of any adequate investigations into the use of the voice tremor phenomenon in a stress situation, the use of micro-tremor in the analysis of stress may still prove to be a useful approach.

General Spectrographic Measurements. These methods attempt to quantify sound spectrograms either by visual inspection, or by direct measurement. Such methods can be effective in demonstrating changes in the voice, but it is difficult to obtain precise measures. The most important spectrographic analyses have used either wide band filters (200-400 Hz bandwidth) to emphasise the formant resonances in the speech spectrogram, or narrow band filters (less than 50 Hz bandwidth) to highlight the harmonic structure due to fundamental frequency.

Kuroda et al (Ref 13) have defined a quantity from the narrow band spectrogram known as the vibration space shift ratio (VSSR). This is simply derived from measurement of the frequency band spacing during voiced speech, and relates to the relative changes in fundamental frequency between normal and high stress situations. Thus if in the normal situation, frequency band spacing is given by SVS, and in the high stress situation, by EVS, then

$$\text{VSSR} = \frac{\text{SVS} - \text{EVS}}{\text{SVS}} \times 100\%$$

Real situations in which military pilots found themselves in difficulties were examined. Highly significant increases in fundamental frequency were reflected in the VSSR, but each case represented a catastrophic situation and three are known to have resulted in a fatal accident. Generally, in such situations, machine analysis is not necessary to demonstrate the gross increases in fundamental frequency attributed to both intense fear and concomitant increases in voice loudness. In the more commonly encountered high workload or stressful situation, changes in voice parameters would be expected to be much less dramatic, and only then would it be necessary to use some form of machine analysis.

More general evaluations of the way stress may appear in the spectrogram have been carried out by Williams, Stevens et al (Ref 6, 7, 18). Some early studies used data, and produced results, which were very similar to those detailed above, although the fundamental frequency contour was also deemed to be of importance. More comprehensive studies attempted to extract as much information as possible from the spectrogram, largely by inspection. For instance, irregular structure in the second and third formant regions of a wide band spectrogram is thought to reflect a non-stable glottal waveform. The results of these studies have been summarised by the changes in voice attributable to four emotions, namely anger, fear, neutral and sorrow. These four emotions, in that order, tended to produce a fundamental frequency which decreased in magnitude and range. Irregular glottal pulses were often seen in the anger and sorrow situations, while unusual pitch contours were characteristic of the fear situation. Changes were also noted in the syllabic rate and duration of utterances. However, it should be noted that the majority of these results were obtained in the laboratory situation. Two methods have been used. Early methods attempted to induce stress using an arithmetic task (Ref 6), but this is likely to produce failure stress as well as task induced stress. Wide intersubject variability was observed. The second method used actors, and the majority of the above results were obtained in this situation. Between the different emotions, the actors were able to produce clear changes in their speech, but the application of these results to the real situation is open to question. This is particularly true of the flying task, where the range of emotions is not directly applicable, and where again, changes in the voice characteristics of highly trained pilots may be expected to be subtle in all but the most extreme situation.

Average Spectrum Measurements. A second example of stress analysis methods which makes use of spectral information uses the average spectrum. This is the spectrum of a complete utterance, and may involve a single word or a longer phrase. Changes in fundamental and formant frequencies during the utterance give pitch and formant peaks a wider and flatter appearance in the average spectrum, and reflect the overall characteristics of the voice during the complete utterance.

Tishchenko (Ref 15) has suggested that formant frequencies tend to change in the stressed situation, and that spectral intensities within these bands also change. This led to the definition of the formant momentum, which is the product of a formant frequency and its intensity. The data in Tishchenko's study consisted of speech from 23 students before, during and after their first parachute jump. All of the spectral analysis methods were analogue. Generally, the first formant momentum increased in the stress situation, and the second and third moments usually decreased although greater variability was observed. This behaviour was explained physiologically, but account had to be taken of the different vowel sounds present in the single words which were analysed. Thus shifts in formants due to different vowels could augment or reduce apparent shifts due to stress.

Popov, Simonov, Frolov et al have attempted to analyse stress and emotions using spectral methods (Ref 16, 17, 24). Early work was directed towards the measurement of changes in the average formant structure of single words (Ref 16). Results similar to those suggested above by Tishchenko were reported. The data were obtained from actors, but further studies used speech from the cosmonauts in the Voskhod 2 spacecraft. Again a centroid spectrum method was used to give an indication of relative shifts in formant peaks. Analogue spectral techniques defined the average spectrum centroid as

$$f_I = \sum_{i=1}^{u} f_i P_i / \sum_{i=1}^{u} P_i$$

where the f_i are the filter tuning frequencies and the P_i are the average power outputs of the filters, measured over the effective time of output for each filter. Choice of U appeared to be empirical, but significant relationships were established between relative changes in the centroid and heart rate, during various stages of the space flight. This, together with a knowledge of the cosmonauts' tasks at each flight stage, suggested that changes in f_I could reflect a stressful situation. Further studies have analysed the envelope shape of the output of each of the bandpass filters. Specifically, the time integral of the output envelope of each bandpass filter is calculated as

$$\alpha_i = \int_0^T U_i(t)\ dt$$

where T is the analysis period and the U_i are the envelope shapes. It is suggested that empirical combinations of the α_i can actually distinguish between different types of emotion, labelled as fear, anxiety, joy and delight (Ref 16-17). However, reasons for the choice of α_i combinations are not explained.

Williams and Stevens (Ref 7) have described similar methods in which they have analysed the average spectra from several seconds of speech. Their findings are merely consistent with increases in speech loudness in the "anger" situation, and a decrease in speech loudness in the "sorrow" situation. The data were derived from actors' speech.

Direct Measurements on the Speech Waveform. Recent studies by Simonov et al (Ref 23) have suggested that crude measures of fundamental frequency (designated by F_{OT}) and first formant frequency (designated by F_O) may be used to discern emotional states. Apparently, these parameters were extracted directly from the speech waveform, but the analysis methods were not described in any detail. However, variations in each of the parameters, approaching 100%, were reported. The use of a discriminant function in the F_{OT} - f_o plane was therefore suggested to differentiate between so called states of rest and emotion. Again, the main bulk of the work was performed with data derived from actors' speech, although the validity of the method was supposedly confirmed using speech obtained from amateur parachute jumpers. To the amateur, a parachute jump clearly presents a highly stressful situation, but no mention was made of possible physical stress interactions.

We have performed similar experiments with the voice of a commercial airline pilot (Ref 25). Cepstrum methods were used to obtain fundamental frequency estimates and to smooth log magnitude spectra, from which formant information could be extracted (cepstrum methodologies are described in the next section). The data consisted of 22 landings into various international airports. For each landing, the baseline or unstressed fundamental frequency and first formant parameters were obtained from about 30 seconds of speech at the top of descent. Parameters in a stressed situation, as indicated by an increase in heart rate, were obtained from about 30 seconds of speech taken around the touchdown instant. Thus for each landing, specified by the index i, the data yielded four parameters: $\bar{f}_{oi}|_u$ and $\bar{f}_{oi}|_s$ are respectively the unstressed and stressed mean fundamental frequencies, and $F_{1i}|_u$ and $\bar{F}_{1i}|_s$ are respectively the unstressed and stressed mean formant frequencies. Normalisation of the data was carried out in terms of the overall mean parameter values derived from the unstressed data in all 22 landings. Thus

$$f_{omean} = \frac{1}{22}\sum_{i=1}^{22} \bar{f}_{oi}|_u$$

$$F_{1mean} = \frac{1}{22}\sum_{i=1}^{22} \bar{F}_{1i}|_u$$

and so $\hat{f}_{oi}|_u = \bar{f}_{oi}|_u \ / \ f_{omean}$

$$\hat{f}_{oi}|_s = \bar{f}_{oi}|_s \ / \ f_{omean}$$

$$\hat{F}_{1i}|_u = \bar{F}_{1i}|_u \ / \ F_{1mean}$$

$$\hat{F}_{1i}|_u = \bar{F}_{1i}|_s \ / \ F_{1mean}$$

where ^ signifies normalized. These data are summarised in Fig 5 which plots parameter variations from the stressed and unstressed centroids in each of the 22 landings. The centroid of the unstressed data lies at the origin, but, it can be seen that the stressed data centroid is shifted to a position representing an increase in first formant frequency, but a decrease in fundamental frequency. The distance between centroids reflects the degree to which stress is manifested in these two speech parameters. An application of the T^2 test to the raw data demonstrates a difference in centroids ($P < 0.0002$), but Fig 5 does suggest that the discriminating power of these two clusters may be somewhat restricted (Ref 26). A section of speech obtained from either the top of descent or just before touchdown, cannot be assigned to the stressed or unstressed group with any degree of certainty, at least not solely on the basis of fundamental and first formant frequency measurements. These conclusions are at variance with those of Simonov et al (Ref 23) who have claimed much greater stress induced changes in their speech parameters.

It is apparent that some form of measure on fundamental frequency and its variations is essential in an analysis of stressful situations. Various measures on the formants, in particular the formant frequencies, appear to be quite promising.

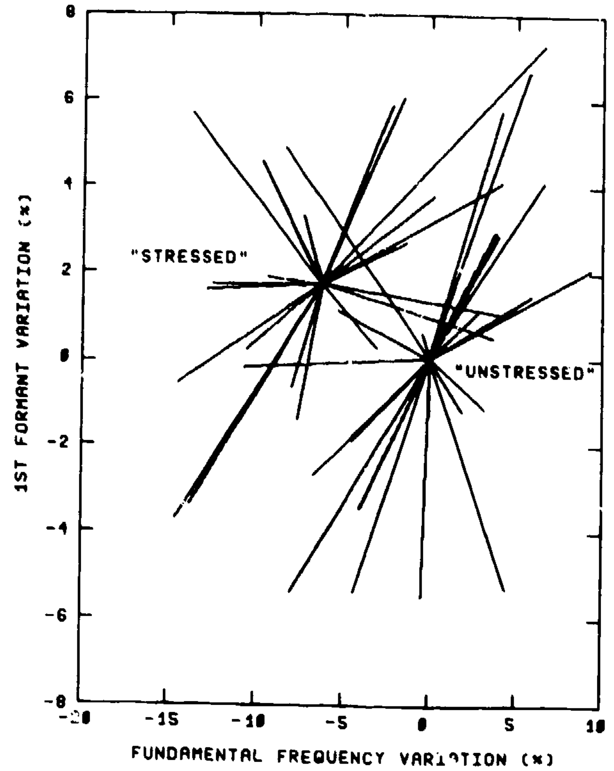


Fig 5
Summary of fundamental and formant frequency variations during 22 commercial landings

SPEECH PROCESSING TECHNIQUES

Speech analysis literature provides a wealth of information covering many areas of application, all of which are based on a requirement to reduce a speech signal to a concise set of parameters. The applications fall into two categories.

1. Reduction of bandwidth requirements in speech communication channels and the automatic machine generation of speech. Neither of these applications is relevant to the current discussion.

2. Speech recognition applications. This broad area requires recourse to statistical and pattern recognition techniques to classify the sets of speech parameters. Speech recognition can mean the extraction of phonetic or semantic information, and the methods which have been developed in this area are often based on prototype or template speech parameters (Ref 27). In the context of stress and high workload analysis, the techniques of speech recognition which are of greater interest are those aimed at identifying the speaker rather than his speech. Considerable effort has been expended in developing means of assigning the voice parameters of an arbitrary speaker to a specific "library" parameter group. It is possible to identify particular speakers from relatively large populations if a section of their speech is available for reference purposes. The relevance of these methods is obvious, and it may be possible to assign short sections of speech from a given speaker to known stressed or unstressed parameter groups. It may also be possible to develop a system using several levels of stress, rather than a stressed-unstressed binary quantisation.

A description of all the available processing techniques cannot be attempted but two techniques which are felt to be directly relevant in the analysis of stress and workload will be considered. The use of cepstrum techniques is common and so the implications of these methods will be considered in some detail.

Cepstrum Techniques. Cepstrum analysis is a powerful methodology which may be used to analyse a voiced speech signal by separating out the contribution due to the glottal pulse and the contribution due to the formant filters. It is possible to identify voiced and unvoiced intervals, and within the voiced intervals to obtain estimates of fundamental frequency. Further, the cepstrum technique can provide smoothed spectral estimates from which formant information can be obtained (Ref 9, 10). Using the model presented in Fig 2 the voiced speech output signal may be assumed to be a convolution of the vocal source impulse train with the impulse responses of the various filters in the system. Thus, denoting convolution with a *,

$$x(t) = p(t) * s(t) * F(t) * r(t)$$

where $x(t)$ is the speech output
$p(t)$ is the impulse train
$s(t)$ is the impulse response of the source filter
$F(t)$ is the combined impulse response of the formant filters
and $r(t)$ is the impulse response of the radiation load filter.

Combining the effects of the source and radiation load so that the vocal source output assumes the form

$$S(t) = p(t) * s(t) * r(t)$$

then $x(t) = S(t) * F(t)$

or equivalently, since convolution in the time domain is identical with multiplication in the frequency domain,

$$X(\omega) = S(\omega) . F(\omega)$$

where $X(\omega)$ is the speech magnitude spectrum.

A logarithmic transform has the effect of separating the elements of $X(\omega)$ into additive components, ie

$$Ln\{X(\omega)\} = Ln\{S(\omega) . F(\omega)\} = Ln\{S(\omega)\} + Ln\{F(\omega)\}$$

$Ln\{X(\omega)\}$ has the appearance of an undulating function representing formant structure with a superimposed "high frequency" ripple representing the harmonic structure of the vocal source spectrum. The additive components in $Ln\{X(\omega)\}$ are maintained during inverse frequency transformation which results in the so called cepstrum. Clearly, the harmonic structure in the log magnitude spectrum manifests itself as a sharp peak in the cepstrum from which pitch period may be determined. There are available, several efficient algorithms which implement cepstral pitch peak picking, and we have developed an algorithm based on a design by Noll (Ref 28), which has proved to be very useful.

This technique, then, is a relatively simple method of measuring fundamental frequency, and is based on the harmonic structure of a log magnitude spectrum. As a consequence, the fundamental frequency component need not be present in the signal being analysed. Further, our experience has shown that the method works well, even in the presence of considerable noise, for instance, with a signal to noise ratio as low as 5 dB, (as defined only during voiced intervals). In this context, noise refers to the acoustic noise in the cockpit environment as well as to any electrical noise introduced by the communication and recording equipment.

It is of interest to compare the cepstrum method with other simple pitch extraction routines. McGonegal <u>et al</u> (Ref 30) have evaluated cepstrum methods together with low pass filtering and autocorrelation techniques. The autocorrelation function is quite similar to the cepstrum except that in the latter, pitch peaks are more pronounced due to the logarithmic transform in the spectrum. With the exception of identifying voiced-unvoiced transitions, the three pitch extraction methods mentioned above were shown to be quite similar in operation.

Cepstrum analysis however, has other advantages when searching for parameters to characterise the spectrum. Since the log magnitude spectrum and cepstrum are Fourier transform pairs, the low order coefficients in the cepstrum contain spectral envelope shape information. This observation provides us with two alternative methods for obtaining spectral information.

1. The low order cepstrum coefficients may be used directly as parameters which classify the speech spectrum .

2. The cepstrum may be short time filtered and transformed back into the log magnitude spectrum. Formant picking algorithms may then be implemented to characterise the spectrum.

The first method is computationally faster, and has found extensive use in talker identification applications. However, the second method is less prone to corruption due to noise in the original speech waveform and is physically more meaningful. It is suggested that in the current application, the second method offers the more viable proposition. Within the framework of cepstrum analysis, there are several methods of extracting formants. Schafer and Rabiner (Ref 10) have provided a robust method (peak picking) which makes full use of empirical data. Alternatively, Olive (Ref 29) uses the model of speech production in an iterative spectrum matching technique (analysis by synthesis). Both of these methods make use of amplitude information, but this is not practical in the current application. We have had some success using an algorithm based on Schafer and Rabiner's design. The algorithm disregards amplitude information except for relative changes within specific formant ranges, and current formant peak picking decisions are, in part, based on previous decisions.

At this stage we must consider the choice of analysis interval, that is the length of the speech epoch used to obtain fundamental and formant frequency estimates. Since these parameters will vary during an utterance, the analysis interval should be arbitrarily short. In practice of course, a compromise is necessary. At least four pitch periods are desirable to obtain a strong peak in the cepstrum, but within this time it is quite likely that one or more of the formant peaks will have moved, producing a smearing effect in the spectral envelope. Generally speaking, the analysis interval is chosen to contain up to four pitch periods (20-40 mS), but successive intervals are overlapped, to have centres which may be only 10 mS apart. Individual fundamental and formant frequency estimates can be used to form contours or profiles covering a complete utterance.

It is usual to implement cepstrum analysis using fast Fourier transform (FFT) methods, either in hardware or software. An important problem which is closely related to the choice of analysis interval concerns the choice of sampling rate and FFT transform size. Assuming formant information is required, a minimum sampling rate of 8 kHz is desirable: 10 kHz is more usual. For fundamental frequency extraction alone, lower sampling rates may be used, but this will result in significant quantisation error. It can be shown that time resolution in the cepstrum is given by

$$T_{CR} = 1/F_S$$

where F_S is the sampling frequency.

Consider a pitch peak at the nth cepstrum coefficient. Fundamental frequency is then given by

$$fo(n) = 1/(nT_{CR})$$

Resolution in fo(n) is inversely dependont upon n and so for a given sampling frequency, the maximum quantisation error increases as fo(n) increases. This is illustrated in Fig 6 which plots cepstrally derived fundamental frequency against the maximum quantisation error, ΔQ, at that frequency. At the nth cepstrum coefficient the quantisation error is defined as

$$\Delta Q_n = \frac{fo(n+1) - fo(n-1)}{2} = \frac{1}{(n^2-1)\,T_{CR}}$$

Intuitively, for the expected changes in fo induced by stress and high workload situations, ΔQ_n should not exceed 2 Hz. Thus if fo does not exceed 150 Hz, a minimum sampling rate of 8192 Hz is sufficient.

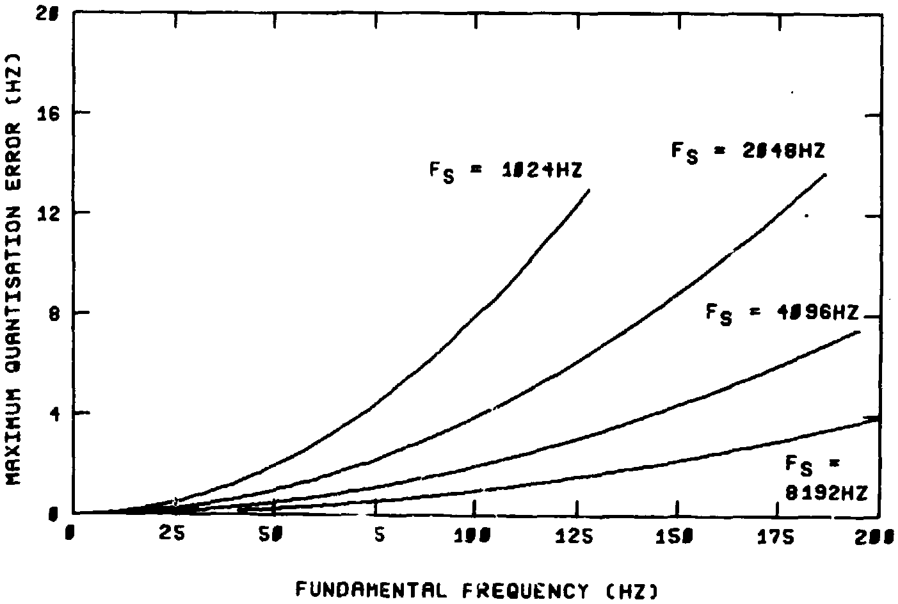


Fig 6
Effect of sampling rate on fundamental frequency resolution

Given a suitable sampling rate, choice of transform size is restricted by the analysis interval requirement. But, if formant picking is to be implemented, a large transform size is desirable to give good resolution in the spectrum and to avoid aliasing problems in the cepstrum.

Now $B_s = F_S/S$

where B_s is the spectral resolution

and S is the transform size.

Given an 8192 Hz sampling rate, a minimum transform size of 1024 points should be used. This implies an analysis interval of 0.125 seconds. It is therefore usual to pad the actual analysis interval with zeros for transformation purposes. A schematic illustration of the complete cepstrum analysis procedure is given in Fig 7.

Linear Predictive Coding. Linear predictive coding is a form of inverse filtering which models the speech waveform itself rather than various aspects of the speech spectrum (Ref 11, 31). The contribution of the vocal source and vocal tract to the speech signal are not separated out and it is possible to track rapidly changing speech processes which may be lost in the relatively long analysis intervals associated with Fourier methods.

In essence, during the segment of speech to be analysed, the nth speech sample, S_n, is given as a weighted sum of the previous p values.

$$S_n = \sum_{k=1}^{p} a_k\, S_{n-k}$$

The weighting coefficients, a_k, can be obtained by first calculating the prediction error, E_n, as

$$E_n = S_n - \hat{S}_n = S_n - \sum_{k=1}^{p} a_k\, S_{n-k}$$

where S_n is the value of a speech sample and $\hat{S}_n$ is its predicted value. E_n^2 is then averaged over all n in the current speech segment to form a mean square prediction error which is minimised by choice of the a_k. The number of coefficients needed to represent a speech segment is given by p, and

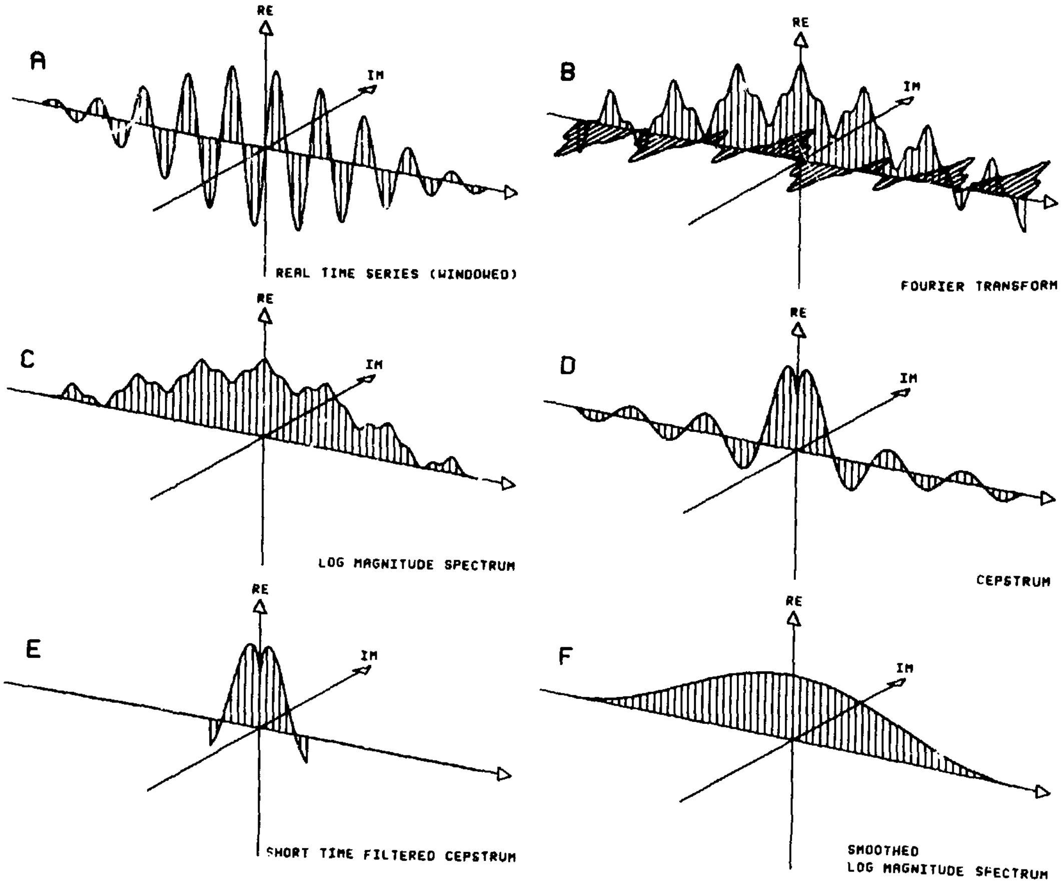


Fig 7
Stylised representation of the stages in a cepstrum analysis

depends on the model chosen to represent the vocal source and tract. It can be shown that twelve coefficients are generally adequate. Thus for a speech signal digitized at 10 kHz, the 100 samples taken over a 10 mSec analysis interval, may be represented by just fourteen parameters, that is twelve weighting coefficients and a pitch period, together with a binary voiced-unvoiced decision.

Fundamental frequency estimates are obtained easily with this method since it can be shown that the prediction error, E_n, is a maximum at the start of each pitch period (Ref 11). A simple peak picking procedure may be used on the E_n series to identify the points in the speech time series at which a pitch impulse occurs. Peak picking is independent of the analysis interval and so very short intervals (as low as 5 mS) may be used, even when fundamental frequency information is required. This offers distinct advantages over cepstrum methods which can only evaluate an average fundamental frequency over the duration of a much longer analysis interval. Furthermore, since predictive coding is a time domain method, it is significantly faster than corresponding frequency domain methods. However, when using predictive coding for pitch extraction it is desirable that the pitch fundamental be present in the digitized speech signal. This, together with the poor quality speech in existing communication channels, makes linear prediction less attractive in the current application.

The advantage of using very short analysis intervals is of more interest if we consider formant extraction. It can be shown that the weighting coefficients, a_k, define the poles in the vocal tract transfer function, and it is possible to obtain both formant frequencies and formant bandwidths. However the problems associated with poor speech quality will again be an overriding factor.

STATISTICAL TECHNIQUES

In the previous section we have shown the way in which speech analysis allows parameters such as fundamental and formant frequency estimates to be derived during a short time period, and, it has been suggested that time series profiles of such parameters may indicate stress or high workload. It is necessary to establish the validity or otherwise of this hypothesis, and so we will briefly consider some of the statistical techniques which may be used to classify or to group series of speech parameters.

The first question concerns the length and nature of the speech epoch to be analysed. To some extent this will be dependent upon the type of the statistical analysis. For example, the examination of single words or single phonems is only feasible under a limited set of conditions. The same phonem or word must be chosen for analysis from the different stages of a flight profile, and this is particularly important in the case of an analysis attempting to use formant information. With regard to fundamental frequency, even though it is expected to remain essentially constant, our experience has suggested that in the course of a single word lasting less than one second, variations due to intonation mask any possible change due to the workload situation. This problem may be partially overcome by restricting the choice of phonem or word to those which may be obtained from similar circumstances during the flight profile. In this respect, call-signs represent useful information, although the automatic and often emotionless manner in which they are uttered may or may not be an advantage. The analysis of single word call-signs is being actively pursued.

Problems due to intonation, varying formant structure or simply random variations, may be overcome by analysing longer sections of speech and averaging the resulting parameters. Although a simple averaging procedure is a valid approach, any other measure which classifies the shape of a parameter profile should be considered. For example, moments about a fundamental frequency mean are useful since it is unwise to disregard intonation information completely. When considering such methods, the question of a suitable length for the speech epoch arises. Long term feature averaging experiments (Ref 32) have suggested that a mean fundamental frequency obtained over a 20 second epoch reduces the sample variance due to intonation and random variations to acceptable levels. In this context, the epoch describes 20 seconds of voiced speech which represents a considerably longer section of normal speech. In the current application, it is unlikely that such lengthy sections of speech will be available. It should also be noted that this methodology derives long term average parameters using a short time analysis technique and results similar to those obtained using the spectrographic method described in section 4 are to be expected. It is desirable therefore to use parameters other than the simple average.

The above discussion suggests that the voice may be searched for signs of stress of high workload in terms of single phonems and words such as call-signs, or in terms of the properties of longer sections of speech. In either case, a complete data set may be viewed as a series of vectors, $\underline{X}_n$: n = 1,2 .... m. Each vector represents the p speech parameters which are chosen to characterise a particular situation. Thus

$$\underline{X}_n = [x_{n1}, x_{n2} \ldots x_{np}]$$

In the case of single phonem analysis the components in $\underline{X}$ may represent the elements of a pitch or formant profile, while for longer sections of speech, the components in $\underline{X}$ will represent different properties of the whole speech epoch. The m vectors are obtained at various stages of the flight profile, and it would be hoped that differences in the structure of the vectors reflect changes in the stress and workload situation.

Such data require recourse to multivariate statistical methods. Principal component and factor analyses (Ref 33) are obvious candidates. Such methods reduce the dimensionality of the vectors and as a consequence, can be used to demonstrate possible groupings in the original parameters. If consistent changes can be produced in speech parameters, then well defined stressed and unstressed situations will resolve into two distinct groups in the vector space. Of greater importance however, is the fact that these methods form the basis of techniques such as linear discriminant analysis which effectively optimize the ability to distinguish between different groups of parameter vectors. These methods have proved useful in speaker identification experiments, using parameters extracted from long sections of speech (Ref 34), and are felt to be useful in the current application. For example, the data presented in Fig 5 forms the basis of a discriminant analysis using just two parameters: the inclusion of further parameters which may provide better separation of the centroids is desirable.

Mathematically, discriminant analysis is a powerful tool, but well defined "training" data sets are required. In the current application this implies that for each subject considered, it is necessary to obtain sections of speech in both stressed and unstressed situations. If such data can define distinct groups then it is possible to assign an arbitrary sample of speech to one of the two groups. The reliable collection of training data constitutes a major problem, particularly in the stressed situation which is not easily definable. In this respect, the availability of other physiological data is of some importance, at least during training procedures. Thus for the data presented in Fig 5, the stressed-unstressed decision was based partly on a knowledge of the workload patterns in the flight profiles, but mainly on the measured heart rate patterns during the flight. A simple correlation analysis between physiological data and speech data is also proving valuable in establishing which speech parameters contain useful information.

CONCLUSIONS AND RECOMMENDATIONS

The salient features of this review lead to recommendations, which may constitute a methodology for the investigation of high workload using speech patterns.

1. The aim should be to reduce the dimensionality of speech and provide a succinct description of the data. This must be done in a way which preserves the possible stress or high workload information. Also a statistically robust method must be found to classify the reduced speech data into at least two groups of stressed-unstressed parameters.

2. The nature of the data requires some attention. It is suggested that an attempt to extract relatively simple speech parameters from many flights across several subjects is the most viable approach. In the long term, a complex and sophisticated analysis on a limited set of data may not be profitable.

3. For a given flight profile it is desirable to have a knowledge of the likely stress or high workload patterns, together with some indication of their rates of change. This information will influence the type and length of speech samples which may be used. Thus for rapidly changing high workload profiles, only single words or phonems may be used, but longer sections of speech may be employed for slowly varying stress patterns.

4. Our experience has suggested that no matter what type of speech sample is chosen, a cepstrum analysis technique offers a realistic compromise between the degree of processing power required and the amount of

information preserved in the reduced data. At this stage, more involved processing methods, while possibly being more powerful, are not considered to be worthwhile, and indeed, would not offer such reliable results given the poor quality of the original speech samples.

5. Cepstrum analysis can offer short time smoothed spectrum and formant information together with fundamental frequency information. Pitch and formant profiles over longer periods of time may be easily constructed. Previous research in this area has suggested that such measures are of considerable value.

6. Whatever statistical methods are employed, they must be capable of assigning arbitrary speech data to some point on a stressed-unstressed scale. Our experience has suggested that initially, only a binary quantisation of the scale may be possible. In any event, multivariate methods are necessary and linear discriminant analysis looks very promising.

7. It is unlikely that an absolute estimate of stress or workload can be obtained from a single speech sample in isolation.

8. Simple correlation analysis of speech parameters with physiological data appears to be a realistic means of establishing which parameters will be of use in the long term.

9. Finally, voice micro-tremor has found limited uses in commercial devices, but a rigorous study of its possible usage in the current application should not be neglected.

REFERENCES

1. Nicholson, A.N., L.E. Hill, R.G. Borland & H.M. Ferres. Activity of the nervous system during the let-down, approach and landing: A study of short duration high workload. Aerospace Med., 41. 4, 436-446, 1970.

2. Nicholson, A.N., L.E. Hill, R.G. Borland & W.J. Krzanowski. Influence of workload on the neurological state of a pilot during the approach and landing. Aerospace Med., 44. 2, 146-152, 1973.

3. McFeely, T.E. Pupil diameter and the cross-adaptive critical tracking task; A method of workload measurement. Thesis, Naval Postgraduate School, Monterey, California, 1972.

4. Various. A preliminary study of flight deck workloads in Civil Air Transport Aircraft. Flying Personnel Research Committee Report No FPRC 1240, 1965.

5. Flanagan, J.L. Speech analysis, synthesis and perception, 2nd Edition. New York, Springer Verlag, 1972.

6. Hecker, M.H.L., K.N. Stevens, G. von Bismarck & C.E. Williams. Manifestations of task induced stress in the acoustic speech signal. J. acoust. Soc. Am., 44. 4, 993-1001, 1968.

7. Williams, C.E. & K.N. Stevens. Emotions and speech: Some acoustical correlates. J. acoust. Soc. Am., 52. 4, 1238-1250, 1972.

8. Hess, W.J. A pitch-synchronous digital feature extraction system for phonetic recognition of speech. IEEE Trans. Acoust. Speech and Sig. Proc., ASSP - 24. 1, 14-25, 1976.

9. Oppenheim, A.V. Speech analysis - Synthesis system based on homomorphic filtering. J. acoust. Soc. Am., 45. 2, 458-465, 1969.

10. Schafer, R.W. & L.R. Rabiner. System for automatic formant analysis of voiced speech. J. acoust. Soc. Am., 47. 2, 634-648, 1970.

11. Atal, B.S. & S.L. Hanauer. Speech analysis and synthesis by linear prediction of the speech wave. J. acoust. Soc. Am., 50. 2, 637-655, 1971.

12. Schafer, R.W. A survey of digital speech processing techniques. IEEE Trans. Audio and Electroac., AU 20. 1, 28-35, 1972.

13. Kuroda, I., O. Fujiwara, N. Okamura & N. Utsuki. Method for determining pilot stress through analysis of voice communication. Aviat. Space Environ. Med., 47. 5, 528-533, 1976.

14. Older, H.J. & L.L. Jenney. Psychological stress measurement through voice output analysis. NASA Report No CR-141723, 1975.

15. Tishchenko, A.G. Dynamics of formants in the spectrum of speech as objective indicator of differences between positive and negative emotions. Environ. Space Sciences, 2, 371-375, 1968.

16. Popov, V.A., P.V. Simonov, M.V. Frolov & L.S. Khachatur'yants. The articulatory frequency spectrum as an indicator of the degree and nature of emotional stress in man. NASA Tech. Translation No TT F-13772, 1971.

17. Simonov, P.V. & M.V. Frolov. Utilization of human voice for estimation of man's emotional stress and state of attention. Aerospace Med., 44. 3, 256-258, 1973.

18. Williams, C.E. & K.N. Stevens. On determining the emotional state of pilots during flight: An exploratory study. Aerospace Med., 40. 12, 1369-1372, 1969.

19. Counterintelligence and Security Inc. 'Dektor' Psychological Stress Evaluator. Port Royal Road, Springfield, Virginia.

20. Communication Control Systems Inc. Voice Stress Analyser Mark IX-P. Third Avenue, New York.

21. Inbar, G.F. & G. Eden. Psychological Stress Evaluators: EMG correlation with voice tremor. Biol. Cybernetics, 24, 165-167, 1976.

22. Gallagher, G. Psychological Stress Analyser. Protection of public figures symposium, Fort Belvoir, Va., 1972.

23. Simonov, P.V., M.V. Frolov & V.L. Taubkin. Use of the invariant method of speech analysis to discern the emotional state of announcers. Aviat. Space Environ. Med., 46. 8, 1014-1016, 1975.

24. Luk'yanov, A.N. & M.V. Frolov. Signals of human operator state (Chapter 6: The Speech Signal). NASA Tech. Translation No TT F-609, 1970.

25. Cannings, R., R.G. Borland, L.E. Hill & A.N. Nicholson. Voice analysis and workload during the letdown, approach and landing. Aerospace Med. Association meeting, Washington, 1979.

26. Anderson, T.W. An introduction to multivariate statistical analysis (Chapter 5). New York, Wiley, 1958.

27. White, G.M. & R.B. Neely. Speech recognition experiments with linear prediction, bandpass filtering and dynamic programming. IEEE Trans. Acoust. Speech and Sig. Proc., ASSP-24. 2, 183-188, 1976.

28. Noll, A.M. Cepstrum pitch determination. J. acoust. Soc. Am., 41. 2, 293-309, 1967.

29. Olive, J.P. Automatic formant tracking by a Newton-Raphson technique. J. acoust. Soc. Am., 50. 2, 661-670, 1971.

30. McGonegal, C.A., L.R. Rabiner & A.E. Rosenberg. A Semi-Automatic Pitch Detector (SAPD). IEEE Trans. Acoust. Speech and Sig. Proc., ASSP 23. 6, 570-574, 1975.

31. Makhoul, J. Linear prediction: A tutorial review. Proc. of the IEEE, 63, 561-580, 1975.

32. Markel, J.D., B.T. Oshika & A.H. Gray. Long term feature averaging for speaker recognition. IEEE Trans. Acoust. Speech and Sig. Proc., ASSP 25. 4, 330-337, 1977.

33. Morrison, D.F. Multivariate statistical methods (Chapters 7 and 8). New York, McGraw Hill, 1967.

34. Hunt, M.J., J.W. Yates & J.S. Bridle. Automatic speaker recognition for use over communication channels. Proc. IEEE Int. Conference, ASSP 77. Hartford, Ct., 1977.

AN EXPLORATORY STUDY OF PSYCHOPHYSIOLOGICAL MEASUREMENT AS INDICATORS OF AIR TRAFFIC CONTROL SECTOR WORKLOAD

by

Richard E. McKenzie, Ph.D.
Edward P. Buckley, Ph.D.
Kiriako Sarlanis, M.A.
Federal Aviation Agency
National Aviation Facilites Experimental Center,
Atlantic City, New Jersey

INTRODUCTION

There has been an ever-increasing concern within the Federal Aviation Agency for the possible adverse effects of stress inherent in the character of the work of Air Traffic Control Specialists (ATCS). Someone has characterized the job of the pilot as involving "hours of routine monotony interspersed by moments of sheer terror." Perhaps this is no less true of the job of the controller whose basic task is to maintain an orderly flow of air traffic, maintain the safe separation of enroute and coverging terminal traffic, and to assist the pilot, often under adverse flying conditions.

As pointed out by Dougherty, Trites, and Dille, who compared health information between ATSC and non-ATSC personnel, those who are engaged in this particular occupation, as well as external observers of the job situation, feel that there is inherent stress involved in the work which may have adverse effects (1). These effects undoubtedly involve internal stress factors such as fatigue, aging, and job experience, as well as external factors of potential aircraft conflict, workload, critical incidents, and other aerospace events. The major concern of human engineering has been to develop command and control systems wherein better displays and more functional controls would enable the controller to better perform his demanding task and ultimately render it less stressful. Basic to this concern has been an attempt to define the controller's task and to identify certain aerospace events, such as number of aircraft, aircraft speed, control sector size, etc., which may be crucial factors in the controller's job performance (2, 3). However, such studies have served only to point out that the real need in evaluating the efficiency of control systems, or of the operator himself, is the establishment of relevant criterion measures. Studies in this area, to date, have demonstrated that simple measures of various aerospace events which comprise the controller's workload do not fully relate to the complex stresses that are experienced in the job performance.

Since external job-related measures do not offer satisfactory criteria, we have turned to internal operator-related measures in an effort to determine their usefulness in evaluating the stressors inherent in the work of the ATCS. Therefore, this study was designed to explore the possibility that certain physiological measures could be related to some aspects of the controller's task, namely, workload defined in terms of number of aircraft (traffic density), and the occurrence of aircraft conflicts.

PROCEDURE

Stimulus Materials: Since the research goal was to determine if selected physiological variables were related to controller workload, the stimulus materials were selected to provide two extremes of work level. This was done by simulating a PPI (Plan Position Indicator) display of an enroute sector by means of specials films using the "CODE" (Controller Deicision Evualtion) technqiue developed by one of the authors (4). One film presented a traffic pattern of low density, i.e., few aircraft and few conflicts. The other film presented a high density traffic pattern, i.e., many aircraft, many more conflicts, and higher aircraft speeds. An aircraft conflict is defined in terms of aircraft in flight that approach each other in such a manner so as to violate established separation criteria. This is a potential collision situation.

The problems were approximately 40 minutes in duration. The low density sample had an average of 6.6 aircraft under control at a given time. The average aircraft speed was 470 knots with a range of 380 knots to 550 knots. The number of conflictions occurring during the problem was four.

The high density sample had an average of 19.4 aircraft under control at a given time. The average aircraft speed was 476 knots with a range of 346 knots to 566 knots. The number of conflictions occuring during the problem was 16.

Subjects: Ten subjects, all Air Traffic Control Specialists, were selected. Their chronological age range was from 29 to 46 years with a mean of 35.7. Their experience as controllers ranged from 7 to 18 years with a mean of 11.5.

Instructions: In order to standardize the subject's approach to the experimental task, the following instructions were read to each man individually:

> You are being asked today to take part in an experiment concerning controller workload and certain related physical changes. We have developed a film showing traffic on an enroute sector. The picture will change every six seconds to give an approximation of a six second sweep on a PPI scope. Some of the aircraft develop conflictions in the film. Your task will be to discern when conflictions are developing and to indicate this fact by pressing the button on your left and noting the identities of the aircraft in confliction on the sheet before you. The number of aircraft you will be responsible for may be unusually high and be going unusually fast--but do what you can. There are two sectors displayed. You are responsible for the conflict detection task for both sectors. Your separation standard is 5 miles and 1,000 feet for the total area. YOUR PERFORMANCE IS SCORED ON HOW ACCURATELY YOU PERFORM. Avoid reporting too soon as this may be only a potential conflict. Avoid reporting too late--prevention action could not be

taken. Both of these factors will be considered errors in determining your score. The records of your performance are for research purposes only and will not be divulged to anyone for any other purpose.

Note that you are asked for two things: (a) to press the button on your left at the time which you would normally instruct one of the pilots to take preventive action, (b) note the identities of the aircraft involved.

All aircraft in the system are identified by alphanumerics. Information on altitude and changes in altitude is also given. Frame 6 (sample display) shows how this information will be given throughout this film. The scale of the map is in the upper left.

Are there any questions?

Test Schedule: All subjects were scheduled for one session during which they monitored the display and performed the required task of conflict identification for both the high and low density traffic patterns. The order of presentation of either the high or low density stimulus was counterbalanced across subjects to rule out order effects. Each traffic pattern was monitored for 40 minutes. Thus, the subjects were instructed, instrumented, calibrated, and then monitored either the high or low film for 40 minutes. Then they had a 10-minute break during which time the film was changed. Then followed another 40-minute session using the alternate film.

Physiological Measurements: Two physiologic measures were selected as dependent variables: Heart Rate and Psychogalvanic Skin Response, also called the Galvanic Skin Response (GSR). Heart rate was recorded via two electrodes attached at conventional locations on the chest approximately 6.5 cm above and below the left nipple on the mid-clavicular line.

GSR was recorded from electrodes attached on each hand in the fleshy area commonly called the "heel." Anatomatically, this area can be described as laying over the fourth metacarpal bone and about 2 cm. to the ulnar side of the palmar aponeurosis. Prior experience has demonstrated that excellent GSR responses can be obtained at this site without the movement artifacts usually associated with a central palmar location. In fact, our subjects were above to write without producing noticable artifacts.

The electrodes were of local manufacture using silver-silver chloride material 1½ cm. in diameter and mounted in a plastic cup measuring approximately 3 cm. in diameter. The electrolyte used was EKG-SOL (Beck-Lee Corporation). The sites were prepared by sponging with acetone and the electrodes, filled with the electrolyte, were attached by Eastman 910 adhesive.

Leads from this sensor were fed into appropriate couplers of an E. and M. Physiograph. Both heart rate and GSR were condenser-coupled. Read-out was continuous at 120 mm. per minute.

Subject calibration for GSR determinations was done by the "sniff" method. Prior to starting the experimental run, the subjects were requested to sniff (a rapid inhalation through the nose) at ½ minute intervals, the GSR amplifier was adjusted to yield a 20 mm. per excursion. When the subject's "sniff" response stabilized at this level the session began. In order to maintain the GSR at a "standard" level throughout the experimental session, the subjects were requested to give a sniff response every minute. By this method, each subject's GSR could be calibrated and maintained throughout the experiment.

Data Reduction: Heart rate was scored on a minute-to-minute basis and an average determined for each experimental session. GSR was evaluated in two ways. First, total GSR frequency was determined in terms of a "standard unit." This standard unit was arbitrarily selected as a 5 x 5 mm. square area. All GSR responses which were less than this size area were ignored. The records were then hand scored in terms of this standard unit. In order to check on the accuracy of the hand scoring, GSR's from each record were randomly selected and the area scored by means of a planimeter (Keuffel and Esser 4211). By this means the hand scoring was determined to be in error less than one percent. Hand scoring for area is a laborious procedure but, in the absence of an electronic integrator, reasonable accuracy can be obtained, although we are not advocating the procedure.

Statistical Analyses: The means and medians for the high and low density situations were computed for all four measures. Both parametric and non-parametric tests for the statistical significance of differences were done for all four measures. The matched pairs 't' test was the parametric test used. The arcsin transformation was used before applying the 't' test to the percentages of confliction detection. The sign test was the non-parametric test used. One tailed tests were used in all cases.

RESULTS

Establishment of Difference Between Traffic Samples: The confliction detection performance is shown in Table I. That there was a very significant difference in the two traffic samples shown is indicated in the significant difference in the confliction detection performance. This establishes the fact that we were in fact dealing with traffic samples which were markedly different in difficulty.

The fact that some confliction detections were missed in the heavier traffic sample shows that it was probably unrealistically difficult, in accordance with the rationale stated earlier for this pilot-study, of using two markedly different conditions to examine the physiological measures. In addition, the fact that there were considerable overwrites among the alphanumerics on the film undoubtedly markedly increased the difficulty of detections. In general, then, the reader is cautioned not to regard these percentages of conflict detections as operationally valid but rather to keep in mind the rationale for this test as an exploratory study.

TABLE I

CONFLICTION DETECTION IN LOW TRAFFIC DENSITY AND HIGH TRAFFIC DENSITY SITUATION FILMS

Subject	Low Density (%)	High Density (%)
1	100	56
2	100	75
3	100	62
4	100	44
5	100	75
6	75	50
7	100	69
8	100	69
9	75	69
10	100	69
Mean	95	64
Median	95	69

Sign test p = .001 t (transformed data) p = .0005

t (untransformed data) p = .0005

Evaluation of Physiological Measures: The major purpose of the study was to see whether certain physiological discriminated between two traffic samples of different difficulty, i.e., whether they would reflect workload.

Tables II, III, and IV, present the data from three measures: Heart rate, GSR frequency, and GSR area, respectively. It is clear from the tables that the best discriminator is GSR area; that the GSR frequency measure is of moderate definitiveness as a discriminator; and that heart rate is least effective, although discriminating.

TABLE II

HEART RATE DURING LOW TRAFFIC DENSITY AND HIGH TRAFFIC DENSITY SITUATION FILMS

Subject	Low Density	High Density
1	96.68	99.35
2	72.48	73.39
3	90.88	109.37
4	90.05	85.64
5	99.40	105.23
6	105.88	96.58
7	63.49	75.51
8	79.47	79.86
9	65.03	67.66
10	88.09	102.99
Mean	85.14	89.56
Median	89.07	91.11

Sign test p = .055 t test p = .10 ---- .05

TABLE III

GSR FREQUENCY DURING LOW TRAFFIC DENSITY AND HIGH TRAFFIC DENSITY SITUATION FILMS

Subject	Low Density	High Density
1	107	110
2	64	72
3	36	75
4	46	112
5	51	99
6	77	76
7	59	82
8	109	92
9	51	97
10	71	128
Mean	67.1	94.3
Median	61.5	94.5

Sign test p = .055 t test p = .01 ---- .005

TABLE IV

GSR AREA DURING LOW TRAFFIC DENSITY AND HIGH TRAFFIC DENSITY SITUATION FILMS

Subject	Low Density	High Density
1	198	433
2	134	279
3	119	207
4	88	296
5	146	199
6	221	478
7	150	206
8	197	234
9	164	272
10	196	324
Mean	161.3	292.8
Median	157.0	275.5

Sign test p = .001 t test p = .0005

This pilot study, then, within its limitations, has accomplished its purpose of determining whether extensive examination of this type of measure was warranted since it has shown that psychophysiological measurements could at least discriminate the human reaction to traffic situation which were known to differ widely in difficulty.

DISCUSSION AND CONCLUSIONS

The reader should be aware, on the one hand, that the traffic situations portrayed here were different in difficulty in the extreme and that these physiological stress measures may not be as successful in discriminating the human effort differences associated with more normal sector-to-sector or hour-to-hour variations in traffic. On the other hand, if further studies confirm the results obtained here, a tool for systems research and development of considerable importance has been found. As only one example of this utility, this methodology could be useful to verify, refine, and improve the recent formulation by

Arad of a mathematical index of the complexity of airspace events (2, 3). Another obvious use is as a criterion for new systems which may have, as one of their values, a reduction in controller effort, fatigue, and stress.

The study has had another outcome, important in the technology of psychophysiological measurement. As previously noted, GSR changes in the subjects were more detectable using variations in measured amplitude area, as compared to frequency of GSR changes. While our methods of evaluating GSR area were laborious due to equipment limitations, the availability of integrating methods for automatically yielding measures of amplitude change should yield important data in evaluating the GSR relative to workload and other stress studies. In this, our work appears to closely parallel a Russian study (5) where Kozarovitskii reports: "In inexperienced airline dispatchers the galvanic skin responses deviated from normal due to fatigue at the end of a working day or under tension. The character of the tracings showed diminishing skin resistance, either an increase or a decrease in amplitude in different individuals, and a decrease in frequency. Distraction of attention tended to lower the skin resistance which depends (on) stimulating and suppressing processes." He also points out that, "Experienced subjects showed fewer sharp variations, less fluctuation in amplitude, a faster fading of the response, and less reaction to the extraneous stimuli."

While this study was of limited scope, it seems that the study of physiological parameters of ATCS workload may yield important criterion measures of external factors of aircraft conflict, task overload, critical incidents, and other aerospace events. Future studies will explore other physiological variables for use as criterion measures, as well as tools for the evaluation of internal stress factors in relation to fatigue, aging, and traffic control experience.

REFERENCES

1. Dougherty, J. D. et. al. Self-Reported Stress-Related Symptoms Among Air Traffic Control Specialists (ATCS) and Non-ATCS Personnel, Aerosp. Med., October 1965.

2. Arad, B., Mayfield, C. E. et. al. Control Load, Control Capacity, and Optimal Sector Design, Franklin Institue Laboratories, Philadelphia, PA and Systems Research and Development Service, Federal Aviation Agency, Atlantic City, NJ, Report No. RD-64-16, December 1963.

3. Jolitz, G. D. Evaluation of a Mathematical Model for Computing Control Load at Air Traffic Control Facilities, Systems Research and Development Service, Federal Aviation Agency, Atlantic City, NJ, Report No. RD-65-69, June 1965.

4. Buckley, E. P. et. al. Pilot Experiments Concerning Air Traffic Control Decision Making, The Franklin Institute, Philadelphia, PA, April 1960.

5. Kozarovitskii, L. B. Dynamics of Skin-Galvanic Reactions in Control Panel Operators During Their Work of Regulating Airplane Traffic, Moscow University, Department of Physiology, Moscow, USSF, Zhurnal Vysshei Nervnoi Deratel' nosti, Vol. 14, May-June 1964, pp. 387-396 (In Russian).

INDIVIDUAL AND SYSTEM PERFORMANCE INDICES FOR THE AIR TRAFFIC CONTROL SYSTEM

by

Edward P. Buckley, Ph.D.*
William F. O'Connor, Ph.D.
Tom Beebe
Federal Aviation Agency
National Aviation Facilities Experimental Center,
Atlantic City, New Jersey

This is a simulation study to examine the relationships between field air traffic controller performance indices and system performance measures. The study encompassed performance criteria developed within two distinct environments, the controller's home facility where he controlled live traffic, and a specially designed microsystem or "one-man ATC system" with simulated traffic. This microsystem simulation was done at the National Aviation Facilities Experimental Center. Thus, the experiment represented a comparative examination of several quantitative measures of system functioning derived from air traffic control simulation and an investigation of these measures as indices for the objective evaluation of the individual air traffic controller.

The initial impetus for this study arose from a concern over the relationship of age and experience to controller proficiency. To a large number of controllers in the field, there appeared a definite trend for older men to be unable to adequately handle the new and complex demands related to the increasing pace in the air traffic control system. While this question of age versus proficiency formed the initial experiment, the basic interest lies in the necessity of the FAA to maintain a highly competent workforce of air traffic control specialists. The central assumption is, that in order to evaluate the effects of age or any other variable upon air traffic controller performance, we must first develope and validate an objective and reliable criterion of performance that has known relationships with controller task functions. Until the establishment of such criteria any question such as the matter of age versus proficiency could only be appraised in terms of indirect or anecdotal measures. Another aspect of this study was that simulation which had been designed and utilized to study procedural and system differences had never been employed for the assessment of performance differences associated with the individual controller. Through the use of simulation in this manner a technique could be developed which would permit the evaluation of each controller operating his test sector as a "micro ATC system" and utilize the system to measure varying system load levels and related controller behavior/efficiency.

Thirty six (36) journeymen enroute air traffic controllers served as subjects having been chosen as a randomized stratified sample of the personnel from four enroute air traffic control centers. The controllers were brought to the NAFEC center for one week in groups of four to receive an orientation to the simulation task including the fictitious geographical area which was to be simulated and the traffic control local procedural rules which were to be in effect in this sector. Each subject performed traffic control during six one-hour runs, with two runs at each of three traffic densities, an experimental protocol designed to produce scores which would measure individual rather than team performance. In addition, each subject was tested on an abbreviated simulation method called CODE, which stands for each subject was tested on an abbreviated simulation method called CODE, which stands for Controller Decision Evaluation. During the main simulation, various performance measures derived from counts or timing of events, for example, the number of aircraft delayed, the delay time, and so forth, were taken. In addition to these system performance measures, stress sensitive measurements of physiological functions were obtained under this dynamic simulation. In addition to the physiological variables of heart rate and galvanic skin response (GSR) measurements, a number of psychological measures were obtained through the use of the 16 PF Test.

The conclusions based upon the results of this project are as follows:

1. The current chronological age of the 36 subjects ranging in age from 31 to 45 years possessed weak, negative relationships with indices of controller proficiency in both field ratings and simulation performance measurements. Age alone, within the range studied, is not a very good performance predictor.

2. Controller age, modified in various ways by experience, does have some effect on performance. This age effect operates in the direction of greater caution and safety, with tendency toward delay of traffic. However, there are wide individual differences within age groups and considerable overlap in proficiency indices between age groups.

3. Current age and age at entrance on duty were highly correlated in this journeymen level group. At the journeyman level, differences apparently due to current age many in fact be due to age at entrance on duty.

4. The personality scale scores based on the 16 PF scales have an unusually large number of statistically significant relationships with controller performance. This suggests that the controllers task which requires sustained performance under complex circumstances, makes such stressful demands as to involve his total personality as well as his skills. The use of these scores as predictors of controller efficiency is not validated. The 16 PF tests reflected that superior controller performance might be linked with the following characteristics: freedom from depression, lack of timidity, socially realistic and relaxed with a relative absence of tenseness or anxiety.

---

* Abstracted with permission of the senior author from final report No. NA 69-40, Federal Aviation Administration, September 1969, by Richard E. McKenzie, Ph.D.

5. In terms of physiological measurements, the controller's heart rate rises in a manner corresponding to increases in the level of traffic and the heart rate levels observed are probably indicative of the stress inducing nature of his task. Unfortunately, the GSR measures did not correspond to the previously reported exploratory study, however, this may be due to the method of evaluation of to the method of GSR collection which was obtained by foot electrodes as opposed to palmar surface electrodes.

6. Simulation system performance measures are reliable and sufficiently precise to measure individual differences in controller proficiency.

7. Simulation measures and field indices of controller performance possess sufficient overlap to establish a meaningful correspondence between the simulation test environment and the live traffic environment. Stimulation technology, then, is capable of providing reliable and objective measurements of controller proficiency.

8. Part measures of the controllers task using the CODE technique appears to be a good objective measure of certain fundamental controller abilities which warrants further development.

9. Results from factor analysis indicate that nine specific system performance criteria are sufficient to describe system functioning over the range of traffic studied.

10. An index which represents the quantification of a trade-off function between volume handling capacity and the occurrence of delays to aircraft called Rdv offers promise as a measure of system load. This index appears suitable for utilization in both the live and simulation system environments for assessing workload differences associated with various sector configurations, staffing patterns, and different geographical areas. The index Rdv is mathematically defined as the correlation between delays and volume in terms of number of aircraft handled at a given traffic level.

Throughout the history of attempts to evaluate individual and system performance, factors simulation techniques and psychological tests have not always proven effective. In this study we see that a simulation system can yield measures of controller proficiency and that at least some psychological test scores can depict superior controller characteristics. Since scores on the 16 PF scales were correlated with both the simulation system performance measures and the field rating measures, further exploration of this test as a predictor of successful controller qualities or as a possible method of evaluating decrements in the performance of career air traffic controllers seems indicated.

WORKLOAD AND STRESS IN AIR TRAFFIC CONTROLLERS

BY

Carl E. Melton
Aviation Physiology Laboratory
Civil Aeromedical Institute
FAA Aeronautical Center
Oklahoma City OK 73125

Abstract

Data collected at 14 air traffic control facilities regarding air traffic controller (ATCS) workload and urinary stress indicator hormone (SIH) excretion is reviewed. The data show a significant relationship between objective workload measures (radio transmission time and traffic counts) and indexes of catecholamine excretion. Mean epinephrine excretion by ATCSs at six air traffic control towers ranging from very low to very high traffic density was significantly (R = 0.96) related to annual traffic counts at those towers. The sympatho-adrenomedullary axis that prepares the organism for "fight or flight" described by W. B. Cannon in 1929 apparently is applicable to ATCSs. The question of underload, optimum load, and overload is discussed.

I. Introduction. The workload experienced by air traffic controllers (ATCS) is difficult to define. One may consider imposed load objectively in terms of numbers of aircraft handled, but the subjective load perceived by the controller may be a greatly different quantity.

Many factors may operate as workload modifiers either making the work easier or more difficult: (1) Type of traffic handled. One aircraft in distress may cause more "work" than all the other traffic being handled. (2) Weather. Controllers' perceived workload always increases when pilots cannot maintain visual separation in instruments meteorological conditions. (3) Equipment outages and malfunctions causing reversion to manual methods of control. (4) Disruption of circadian rhythms caused by rotating shifts, and (5) General physical and emotional condition resulting from a variety of off-duty activities and on-duty problems with management or peers.

It is perceived workload that gives rise to the poorly-defined entity known as stress. Excessive stress has generally been assumed to be a component of air traffic control work and has been legally recognized as such in Public Law 92-297 which provides full retirement for controllers over 50 years of age after 20 years of work controlling air traffic.

Estimates of stress in ATCSs are rendered difficult because of the interaction of off-duty and on-duty experiences. The ATCS undoubtedly brings off-duty problems to work with him and, just as certainly, takes home with him concerns connected with the work place. Thus, a complete representation of stress in ATCSs must integrate all aspects of the ATCS's life.

There has been a strong tendency in the popular press to describe stress in ATCSs in terms of conditions at "hot spot" facilities such as O'Hare and Atlanta Air Traffic Control Towers (ATCT). Generalizations from these descriptions give a skewed idea about stress in the entire population of ATCSs, most of whom work in facilities with far fewer operations.

For the last 10 years this laboratory has carried out studies aimed at providing a general description of stress in ATCSs. These studies have encompassed several variables including numbers of air traffic operations, shift rotation effects, automation, different kinds of air traffic control (ATC) work, and geographical distribution. This report represents an attempt to provide a general concept of stress and workload in ATCSs.

II. Methods. Estimates of stress were derived primarily from urine biochemical analysis for 17-ketogenic steroids (17-KGS), epinephrine (E), and norepinephrine (NE). In most cases values for these stress indicator hormones (SIH) are expressed as creatinine (CR)-based ratios (μg SIH/100 mg CR). Urine analysis for SIH was carried out as previously described (1). Urine collected at field facilities was frozen at the work site; when a sufficient number of specimens had accumulated, they were shipped to the Civil Aeromedical Institute (CAMI) by air freight. Upon receipt the specimens were placed in a freezer where they were kept until analyzed. Specimens were in transit for 3-5 h and detectable thawing did not occur.

Subjects were all volunteer male air traffic control specialists. They were instructed to void and discard urine just prior to retiring the night before a workday. They were then to collect all urine voided until they arose; normally there was only one voiding and that one upon arising. They were then instructed not to collect urine until they arrived at work. At work they were told again to void and discard and to collect in one container (or two, if that one became full) all urine subsequently voided during the workday. ATCSs then repeated this collection regimen for various periods of time depending on the facility being studied. In some studies urine was collected for a whole 5-day workweek; in others, urine was collected for 2 days only because of changeable shift patterns. Each 24-h rest-work period was represented by two specimens.

Urine was collected in cuboidal, plastic 1-quart receptacles containing an excess of dry boric acid as a preservative. When ATCSs delivered the containers to the technical crew the containers were labeled, logged, and frozen.

Workloads were estimated in two ways. One involved the recording of all radio transmissions received by and coming from the subject ATCS. From these recordings total radio transmission time (RTT) was derived by use of voice-actuated relays and digital counters. Workload was also derived from traffic counts.

In order to amalgamate a large volume of biochemical data, a stress index (C_s) was formulated. The details of the index have been published (2). Briefly, the index is based on the idea that the product of the resting and working values for each SIH gives a more realistic view of stress than does the excretion increment (or decrement) from rest to work. However, because the SIHs appear in such unequal quantities in the urine, each individual mean that value (rest and work) is adjusted by dividing it by a grand mean derived for that SIH from all the measurements made on ATCSs in all past studies in this laboratory. This adjustment causes all SIHs to assume equal importance in the calculation of C_s. C_s is the average of indexes calculated for each of the SIHs, c_{st} (17-KGS), c_e (E), and C_{ne} (NE). This index allows different controllers and facilities to be readily compared.

The indexes for each of the SIHs can be presented diagrammatically to show composite stress and the relative contributions of each SIH thereto. The diagram is based on the theorem that the sum of the lengths of internal lines emanating from a common point and perpendicular to the sides of an equilateral triangle is equal to the altitude of the triangle (3,4). The values for c_{st}, c_e, and c_{ne} can be represented as lines originating at a common point and diverging at angles of 120°, the lengths of which are proportional to the values of c_{st}, c_e, and c_{ne}. Lines drawn perpendicular to the free ends of the diverging lines form an equilateral triangle, the area of which is proportional to C_s, the average of c_{st}, c_e, and c_{ne}.

III. Results. Field Experiments. Table 1 shows the correlation between C_s, c_{st}, c_e, c_{ne}, and RTT at Opa Locka (OPF) Air Traffic Control Tower (ATCT) located on a very busy general aviation airport in Greater Miami, Florida. It is apparent that RTT is significantly related to c_e and, with less significance, to c_{ne}. RTT is not significantly related to c_{st}.

Figures 14 show graphically the relationship between stress indexes and RTT at OPF.

Between 1972 and 1974, Los Angeles (LAX) and Bay Area (Oakland (OAK)) Terminal Radar Approach Control (TRACON) facilities were given automated radar terminal system equipment (ARTS-III). This equipment displays aircraft identification and altitude on the radar cathode ray tube, and thus contributes greatly to safety. This equipment was also expected to reduce significantly the workload of radar controllers. There were other changes associated with the new equipment, also. The TRACONs were moved from their towers to separate buildings with adjacent parking lots; the dress code was relaxed; lounge facilities and the general work environment in the control rooms were greatly improved.

Studies were carried out at these two TRACONs prior to (1972) and after (1974) installation of ARTS-III. Table 2 shows changes in stress indexes for individual controllers before and after ARTS-III installation. It is clear that there was a uniform drop in 17KGS and an increase in catecholamine excretion. The workload in terms of number of aircraft worked and number of radio contacts is shown in Table 3.

The traffic count at LAX increased by 3 percent and at OAK by 4 percent from 1972 to 1974. The number of radio contacts increased by 3 percent at LAX and by 1 percent at OAK, while composite stress increased by 25 percent at LAX and 20 percent at OAK. This increase can be seen diagrammatically in Figure 5. Thus, the disproportionate increase in stress, entirely due to elevated catecholamine excretion, is not explained by the objective workload. The explanation most likely lies in work elements not reflected in traffic counts, RTT or number of radio contacts. The new TRACONs had been in use only about 5 months at the times of the second studies. There were still equipment difficulties to be worked out; outages were fairly frequent. Controllers liked the reduction in coordination with other facilities and within the TRACONs; however, the concensus among the controllers was that ARTS-III had not reduced and, in fact, had increased the total workload, primarily because of unfamiliarity with the new equipment.

Catecholamine Excretion and Traffic Count. Because of the demonstrated relationship between E excretion and RTT, annual traffic counts for ATCTs where studies have been conducted were graphed against mean E excretion for controllers at those towers. Such a graph is shown in Figure 6 where annual traffic count is plotted against mean working and resting E excretion for ATCSs at ATCTs ranging from low to high density. The relationship between traffic count and mean E excretion is significant ($R = 0.96$). The working value for O'Hare ATCSs has been displaced to the left to reflect the fact that ORD ATCT was effectively operated as two facilities with separate control positions for the north and south sides of the airport; one side was customarily used for departures and the other side for arrivals. The workload impinging on each ATCS was thus about half of the total airport traffic. When the data point is moved to reflect this division of work, it falls near the line of best fit.

Laboratory Experiments. Because reallife stress arises from a mixture of stressors, the physiological responses to those stressors are difficult to interpret. One cannot separate off-duty experiences from workrelated factors. Therefore, an attempt was made to expose paid experimental subjects to "pure" stressors in the laboratory in order to delineate the specificity of the hormonal response, should there be such.

The subjects (10 young men) were each exposed to a purely physical task with no competitive element (treadmill, 3 miles per hour with no grade) and, on another date, to a purely competitive but nonphysical task ("Pong," a video game based on pingpong). One of the researchers acted as opponent for all subjects; she was an expert and was rarely beaten. Order of presentation of the tasks was balanced; each task was presented in 50-min episodes. In the 10- min following each episode, urine collections were made, rest was allowed and water was imbibed to replace the urinary loss. Urine was analyzed for 17KGS, E, and NE. Values are expressed as the total quantity of each SIH excreted during each 50-min episode. The schedule in each instance was maintained for 3 h

Prior to either experimental exposure, each subject rested for 50-min in the supine position on a cot, ECG electrodes having been previously attached for registration of ambulatory heartrate on small battery-operated ECG tape recorders.

The results of urine and heart rate analyses are shown in Tables 4 and 5. Corresponding episodes of the two tests did not cause significantly different excretion levels in urinary metabolites. Heartrates were significantly higher for treadmill than for Pong tasks. Rest-to-work difference in E excretion was statistically significant for the Pong task, whereas the difference was not significant for the treadmill task. Rest-to-work differences in excretion of NE and 17-KGS were not statistically significant for either task.

IV. Discussion. Data collected over 10 years from several ATC facilities point to catecholamines, principally E, as being a good indicator of the response to an applied workload. The adrenal cortical response (17-KGS) primarily indicates chronic stress arising from unresolved conflicts such as labor management disputes, marital difficulties, financial problems, etc. NE and E usually go in the same direction—seldom does one go up and the other down. It is always difficult to separate the so-called physical and mental stressors in a field study setting. Our laboratory studies indicate strongly that mental activity without significant physical effort engenders a significant output of E above the resting state that does not occur during episodes of purely physical effort.

It thus appears that there is a degree of stressor-response specificity. The large unanswered question relates to the significance of the magnitude of the response. When is a person underloaded, optimally loaded and overloaded? It is also clear that ATCSs at ATCTs with low traffic density have a low E output while ATCSs at high density ATCTs have a relatively greater E output. Obviously, each group of ATCSs is doing at least an adequate job and one cannot say on that basis whether overload, optimal load, or underload is present. However, it is distinctly possible, in view of the known effects of catecholamines on the cardiovascular system, that the cost of adequate performance is greater for the ATCSs at high density facilities than it is for ATCSs at low density ATCTs. Rose (5) has recently shown, as have others in the past (6), that hypertension is more prevalent among ATCSs than among the general population. We have shown that E excretion level is significantly and directly related to heartrate (7). Further, we have shown in a limited group of ATCSs that elevated NE excretion is predictive of later hypertension (8). These data suggest that the cost of adequate performance at a high traffic density ATC facility may result in breakdown of physiological systems.

At the other end of the workload spectrum it is obvious that adequate performance can be maintained without great arousal brought on by high blood levels of catecholamines. In short, it appears probable that arousal necessary to meet workload demand is mediated by sympathoadrenal output of catecholamines. This idea was first put forward by Cannon in 1929 in his description of the "fight or flight" reaction (9) and is applicable to the ATC task.

These data also are consistent with the data reported by Schaad, Gilgen, and Grandjean who showed a statistically significant relationship between urinary catecholamine excretion and level of difficulty of work in European air traffic controllers (10).

REFERENCES

1. Melton, C. E., J. M. McKenzie, B. D. Polis, M. Hoffmann, and J. T. Saldiver, Jr. Physiological Response in Air Traffic Control Personnel: Houston Intercontinental Tower. FAA Office of Aviation Medicine Report No. FAA-AM-73-21, 1973.
2. Melton, C. E., J. M. McKenzie, J. T. Saldivar, and S. M. Hoffmann. Comparison of Opa Locka Tower With Other ATC Facilities by Means of a Biochemical Stress Index. FAA Office of Aviation Medicine Report No. FAA-AM-74-11, 1974.

3. Streng, O. Eine Volkerkarte. Eine graphische Darstellung der bisherigen Isoagglutinationsresultate. Acta Soc. Med. Fenn. Duodecim., Vol 8, pp. 1-17, 1927.

4. Streng, O. Das Isoagglutinationsphanomen vom anthropologischen Gesichtspunkt aus. Acta Pathol. et Microbiol. Scan., Suppl. 5, pp. 59-62, 1930.

5. Rose, R. M., C. D. Jenkins, and M. W. Hurst: Air Traffic Controller Health Change Study. A report to the Federal Aviation Administration on research performed under Contract No. DOT-FA73WA-3211 awarded to Boston University.

6. Dougherty, J.D.: Cardiovascular Findings in Air Traffic Controllers. Aerosp. Med., 38:26-30, 1967.

7. Melton, C. E., J. M. McKenzie, J. T. Saldivar, and M. Hoffmann: Studies on Stress in Aviation Personnel: Analysis and Presentation of Data Derived from a Battery of Measurements. Advisory Group for Aerospace Research and Development, AGARD-CP-180, 1975.

8. Higgins, E. A., M. T. Lategola, and C. E. Melton: Three Reports Relevant to Stress in Aviation Personnel: I. Development of the Aviation Stress Protocol--Simulation and Performance, Physiological and Biochemical Monitoring System: Phase I. II. Assessment of Cardiovascular Function After Exposure to the Aviation Stress Protocol--Simulation. III. The Relationship Between Stress-Related Metabolites and Disqualifying Pathology in Air Traffic Control Personnel. FAA Office of Aviation Medicine Report No. FAA-AM-78-5, 1978.

9. Cannon, W. B.: Bodily Changes in Pain, Hunger, Fear and Rage. C. T. Branford, Boston, MA., 1929.

10. Schaad, R., A. Gilgen, and E. Grandjean: Excretion of Catecholamines in Air Traffic Control Personnel. Schweizerische Medizinische Wochenschrift (Swiss Medical Weekly), 99:889-892, 1969.

TABLE 1. Correlation Coefficients of Stress Indexes and RTT at OPF

INDEX	RTT
C_s	0.64*
c_{st}	0.19
c_e	0.77**
c_{ne}	0.55*

* $P < 0.05$

** $P < 0.01$

TABLE 2. Stress Indices From Individual Controllers Before and After ARTS-III Installation

	c_s		c_{st}		c_e		c_{ne}	
LAX	1	2	1	2	1	2	1	2
Subject								
1	0.90	1.25	0.96	0.34	0.48	1.80	1.25	1.61
2	0.42	0.37	0.64	0.20	0.20	0.50	0.42	0.41
3	0.30	0.36	0.48	0.32	0.09	0.18	0.32	0.57
4	0.48	1.28	0.33	0.21	0.38	1.11	0.72	2.51
5	0.62	0.47	0.55	0.14	0.43	0.67	0.39	0.60
6	0.52	0.47	0.68	0.34	0 20	0.42	0.69	0.65
7	0.42	1.42	0.59	0.33	0.29	0.59	0.38	3.42
OAK	1	2	1	2	1	2	1	2
Subject								
1	0.45	0.68	0.58	0.51	0.53	0.99	0.24	0.54
2	0.73	0.72	0.50	0.27	1.21	1.17	0.50	0.72
3	0.57	0.47	0.56	0.21	0.49	0.59	0.67	0.60
4	0.35	0.71	0.31	0.29	0.29	1.56	0.44	0.28
5	0.46	0.69	0.90	0.37	0.23	1.08	0.25	0.63
6	0.42	0.86	0.59	0.43	0.24	1.27	0.43	0.88
7	0.27	1.36	0.20	0.21	0.28	2.83	0.34	1.05
8	0.18	0.57	1.17	0.05	0.15	1.01	0.22	0.55
9	0.40	0.84	0.73	0.15	0.28	1.81	0.20	0.56
10	0.33	0.47	0.61	0.16	0.22	1.03	0.17	0.22
11	0.22	0.74	0.31	0.10	0.16	1.33	0.20	0.79

TABLE 3. Number of Aircraft Worked and Number of Radio Contacts in a 5-Day Workweek.

FACILITY	NO. AIRCRAFT	NO. CONTACTS	CONTACTS/AIRCRAFT
LAX (1972)	1,803	13,806	7.66
LAX (1974)	1,860	14,210	7.64
OAK (1972)	1,190	8,712	7.32
OAK (1974)	1,238	8,827	7.13

TABLE 4. Comparison of Excretion Values and Heartrates for Pong and Treadmill Tasks*

Task	Total Amounts of Hormones Excreted 17-KGS mg	E ng	NE ng	Heartrate (Beats Per Minute)
Rest (Pong)	0.70	1,237	3,603	64
Rest (T-Mill)	0.67	1,214	4,274	64
P	NS**	NS	NS	NS
Pong 1	0.70	1,619	3,809	73
T-Mill 1	0.59	1,741	4,384	101
P	NS	NS	NS	0.05
Pong 2	0.62	1,720	3,379	73
T-Mill 2	0.67	1,463	3,813	100
P	NS	NS	NS	0.05
Pong 3	0.59	1,750	3,833	70
T-Mill 3	0.58	1,491	3,581	98
P	NS	NS	NS	0.01

* Group Averages
** T-test

TABLE 5. Statistical Significance of Rest-To-Work Differences for the Various Measurements*

TASK	Level of Significance of Difference Between Rest and Task (P**) 17-KGS	E	NE	HEARTRATE
Pong 1	NS	0.01	NS	NS
Pong 2	NS	0.01	NS	NS
Pong 3	NS	0.05	NS	NS
T-Mill 1	NS	NS	NS	0.01
T-Mill 2	NS	NS	NS	0.01
T-Mill 3	NS	NS	NS	0.01

* See Table 4 for actual values.
** Paired t-test.

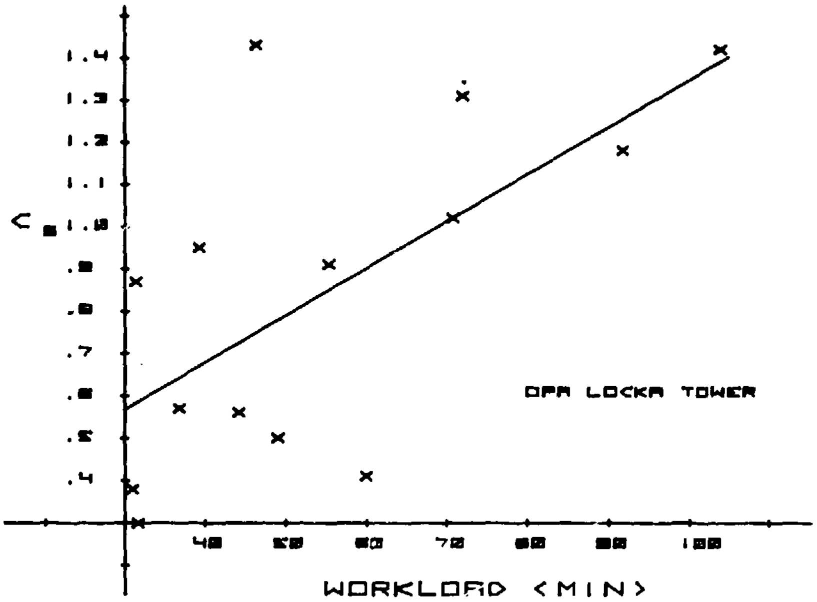


FIGURE 1. Relationship between RTT and C_s at OPF ATCT. The relationship is statistically significant ($p < 0.05$).

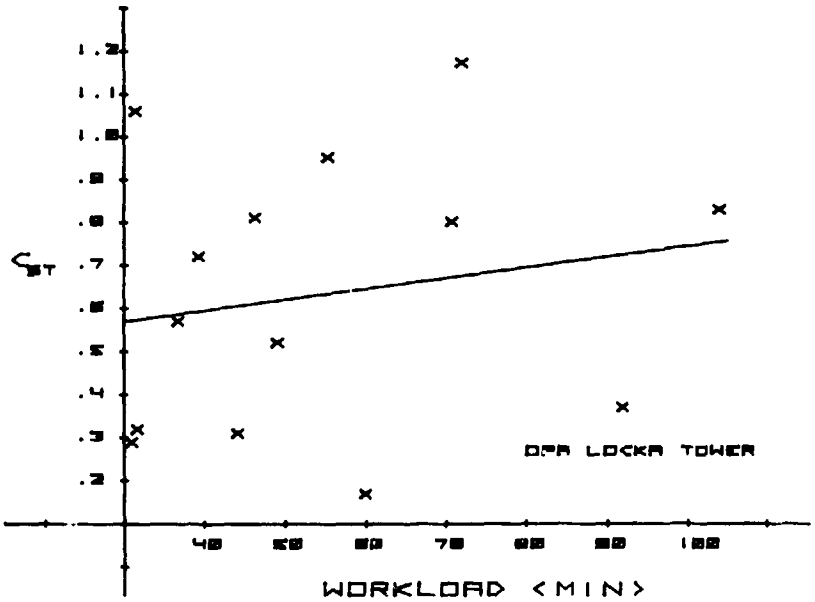


FIGURE 2. Relationship between RTT and c_{st} at OPF ATCT. The relationship is not statistically significant.

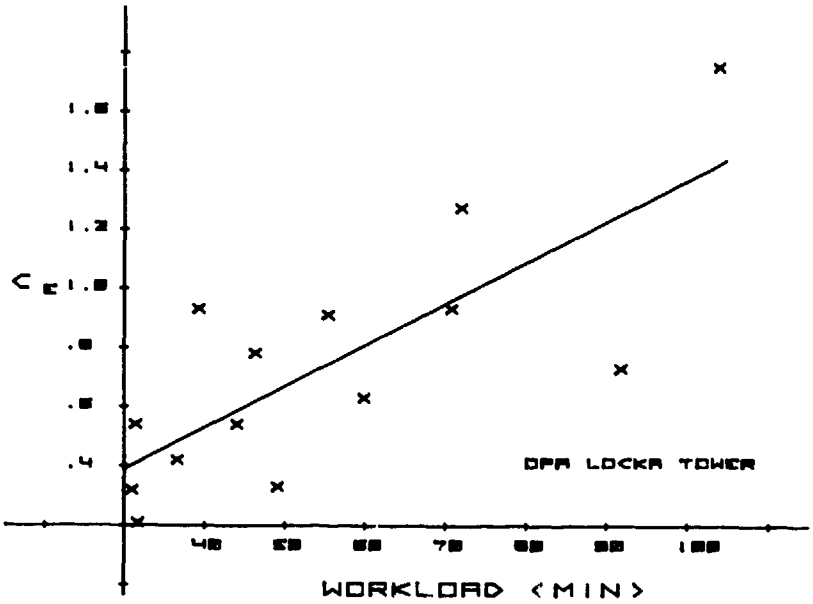


FIGURE 3. Relationship between RTT and c_e at OPF ATCT. The relationship is statistically significant ($p < 0.01$).

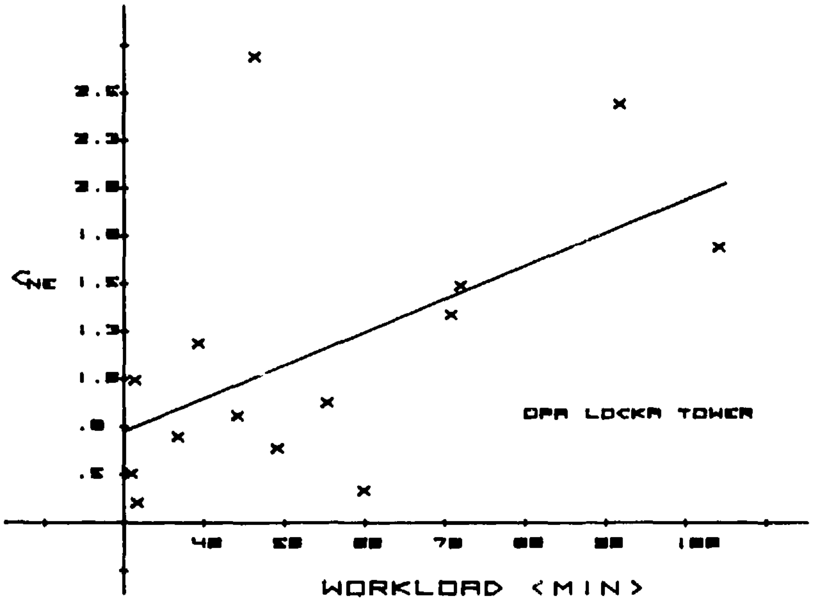


FIGURE 4. Relationship between RTT and c_{ne} at OPF ATCT. The relationship is statistically significant ($p < 0.05$).

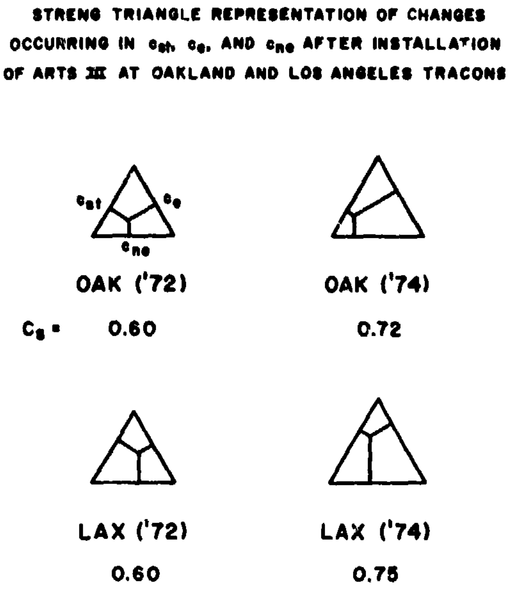


FIGURE 5. Diagrammatic representation of the relationship between c_{st}, c_e, and c_{ne} on Streng's triangle. Comparison of LAX and OAK TRACONs before and after installation of ARTS-III.

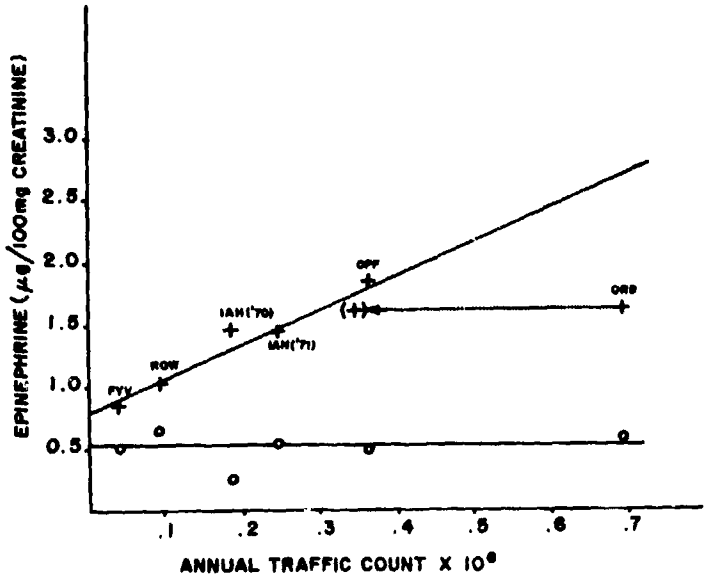


FIGURE 6. Graph of annual traffic count (in millions of operations) vs. mean urinary excretion levels of E of controllers at the various facilities. Crosses represent on-duty excretion levels of E; circles represent corresponding resting levels (ORD graphed at actual traffic count and adjusted value (+) as explained in the text).

ASSESSMENT CORRELATES OF WORKLOAD AND PERFORMANCE

by

Richard E. McKenzie, Ph.D.
Crew Technology Division
USAF School of Aerospace Medicine (AFSC)
Brooks Air Force Base, Texas 78235
USA

INTRODUCTION

A few years ago Dean Chiles (1), while discussing objective methods for developing indices of pilot workload talked about a hypothetical research vehicle having a system with the following capabilities: (1) it would have an exact assignment of the nature and number of pilot duties that could be developed for any given mission; (2) it would be possible to vary those duties in any combination over time; (3) the control and display characteristics of the vehicle could be manipulated at will; (4) precise and reliable quantitative indices of the task demands placed on the pilot by the system would be available for all task elements; (5) precise and reliable quantitative measures of the skill with which the pilot meets these demands would be available; and (6) an adequate criterion measure of total system performance, would be available. If we had such a hypothetical vehicl it is obvious that we would be able to determine the priorities that a pilot assigns certain tasks. We would also be able to assess the attention demands on the pilot by the system, and we would be able to determine which tasks or performance functions are most sensitive to variations and total demand, we might be able to solve some of the problems we have experienced in terms of the human being acting at certain optimal work rates, ignoring certain signals from his display, working at how own pace, failing to pay attention to certain instruments or signals, but in general, always managing to perform at a satisfactory enough level to complete the mission. Perhaps a hypothetical vehicle being a realistic flying system would also help to weed out the variable caused by non-flying laboratory systems where the subject allowed to crash the system because of decremented performance. Whereas in "real world" systems we find that the pilot works harder and harder; performing more and more control responses and movements, but the end result is usually to make a landing that he can walk away from.

Chiles also pointed out that the first and foremost factor to keep in mind in choosing a methodology for assessment is the purpose or goal of the research. Unfortunately, the entire history of assessing workload, performance or stress in the human operator is one of compromise. We have to compromise because of safety, because of operational requirements; we have to devise laboratory-type tasks because the real thing is not available or is unsailable to our measures and often times we must rely upon human beings other than pilots to perform these tasks because of the demands upon pilotage time in the real systems world.

The assessment correlates of workload, performance, and stress can be divided into several areas: those of physiological correlates, psychological correlates, stress correlates, psychophysiologic correlates and finally central nervous system (CNS) correlates. We realize that this is an artificial taxonomy and that many areas of overlap exist; however, we thought we would arbitrarly subsume under the heading psychologic correlates those tasks of vigilance, monitoring, tracking, reaction time, and so forth. It should be noted that we're talking about the operational aspects of psychologic correlates, that many of the psychological tests have been used as selection devices or as measures of skill in order to predict successful training as a pilot or aircrew member. These tests, for the most part, have shown little relationship to the prediction of workload and performance abilities. Many of these tasks have had the disadvantage of single operators looking at single displays and measuring single scores which are often poorly related to the real world of the man-machine interface. Some of these measures are difficult, if not impossible to measure in the operational environment. However, we should keep in mind that it may not always be necessary to measure everything in the operational environment, that is, while the aircrew member is piloting or doing his thing. It may be possible to ascertain whether he is capable of initiating this particular flight-mission and it may be possible to ascertain upon his return from a mission the amount of decrement that resulted by using relatively simple type, field type tasks. Very often in pursuit of psychologic variables we have resorted to the use of subjective evaluation of such factors as fatigue, stress, irritation, etc. Unfortunately, we are finding increasing evidence that certain aspects of these factors may not be amendable to accurate subjective evaluation.

Some 20 years ago this author in his thesis study which had to do with the effect of binaural beats upon performance (2) found out there was a subjectively experienced quality of beats produced externally in an audio mixer as compared with those auditory beats generated centrally in the central nervous system. These two kinds of auditory beat phenomenona were preceived as differentially disruptive by the subjects. In one experiment the subjects said they felt that the stimulus in neither session bothered them in any way especially in terms of their performance on the tests that they were required to do. They ignored the fact that they were not doing any better in consecutive performances and, in fact that they were actually doing somewhat poorer. In another experiment, the subjects invariably reported that the sound of an externally produced beat really bothered them and that they were sure that their performance was effected. This was in spite of the fact that they could observe that they were completing more items and doing better than they had on previous sessions. We were forced to conclude that the neuralmechanism by which binaural beats influenced performance is not open to correct subjective evaluation.

In terms of physiologic correlates of workload and performance we are considering the electromyogram, the electrocardiogram, the measurement of various metabolites in the parotid fluid and urinary tract, etc. Unfortunately, these physiologic correlates, while telling us that a human being has been stressed, do not tell when in the course of time the stress occurred, or what was the nature of the stress. They simply tell us end result of workloads and performance which alter the body's physiology in such a manner that the effects can be measured. Nevertheless, some of these measures do yield interesting correlates of performance.

Stress correlates of workload and performance are somewhat more ambigious since they bridge both the psychologic and physiologic factors. Perhaps, it would be better to consider terms such as environmental and operational stresses. We have to consider the problem of the acuteness of the stress versus chronic stress versus the cumulative effects of stress. Perhaps, we will be able to explore the usefulness of the concept of task-induced stress, where the stress lies clearly within the task in such a way as to present the human operator with a unique situation.

Some of the psychophysiologic correlates that we will consider are the critical-flicker-fusion rate, the psychogalvanic skin response and electro-oculography. These and other measures can be useful in terms of revealing parameters of central nervous system function. We will elaborate considerably in this area because it leads us to the important concept of central nervous system correlates of workload and performance. Nevertheless, psychophysiologic correlates are difficult to relate to actual performance or to workload effects because they may reflect subjective evaluation of task difficulties and they may be related to subjectively calculated probability measures. Nevertheless, if one can tease out these factors one is left with some correlates that may reflect the activity of the central nervous system.

In spite of the various problems of correlates of assessment, we will try to explore some of the various correlates, both old and new, which may offer some help in the quest for measures and assessment of human workload and performance. No attempt is going to be made to make this a global overview. Rather, we have been highly selective in eliminating many measures which appear to offer no fruitful results for the amount of effort expended. We have carefully eliminated any measures which can be regarded as selection tools or measurements of ability, skills and so forth which have little to do with the ultimate question of trying to evaluate or predict human performance in the operational environment.

Psychologic Correlates: The tasks of vigilance, monitoring and tracking seem to have a common root. Most of them involve relatively long times at the task; they involve the detection of a signal or the detection of a nonsignal or nonoccurrence and some form of motor response. Usually, the tasks are of a simple nature, although they may be made more and more complex in terms of additional targets, etc. Such tasks also may be used to evaluate the effects of other kinds of stimuli on vigilance, monitoring or tracking type performance.

In general, we have found that this type of task when considered individually, that is, one task performed by one operator, in one session differ considerably when they are embedded in a multiple task-type simulator or a multiple task paradigm. The vigilance monitoring or tracking task is felt to measure alertness and provides for minimal requirements for intellectual and neuromuscular function. A typical such task is described in the Neptune system (3) wherein the display consists of three meters with a zero centered needle which deflects either left or right and six push buttons, two for each meter. The subject monitors the meters until the needle deflects then he pushes the correct button and the needle returns to the center position. The measure of this performance is response time. Programming of the signals is aperiodic. This is a simple, very undemanding task element, but the behavior it measures is considered important at low levels of arousal. Trumbo (4) points out an experiment in which they had subjects track step-function sequences with six possible target positions. From each position there were two alternative steps, unequally probable and either in the same or opposite directions. The subjects had to anticipate to minimize error, therefore, each step presented them with either an amplitude or a direction prediction problem. The outcome scores which they obtained showed a relationship between input uncertainty and tracking error. Evidence for response strategies or organization came only from continuous records. These records revealed that subjects clearly used different strategies in the two prediction situations, matching event probabilities and predicting direction, but averaging probabilities in predicting amplitude. The importance of this finding is that not only were these strategies unavailable in the outcome scores which are usually measured in terms of response time, but the averaging strategy could not have been identified if subjects had been limited to discrete response alternatives rather than a continuously graded response.

In the measurement of vigilance, monitoring, and tracking activities, usually we have found a number of hybrid tasks put together to evaluate a particular problem. However, in general we find that a nonmechanical electronically driven system such as an oscilloscope provides a display with the most advantages. It is accessible to automated scoring and is readily adaptable to pursuit or compensatory displays and to one or two dimensional courses. Ideally, the response or control apparatus should permit the operator to produce a continuum of graded responses, especially if one is interested in evaluating the aspect of motor skills along with the vigilance or monitoring activity. There are many off-the-shelf function generators to provide a programming system, but the use of a analog computer system would certainly be advisable if one is seriously interested in the determination of probabilities and the ability to vary stimulus organization along many dimensions. Most investigators are not concerned with motor skills for the sake of investigating motor skills themselves, but have used a motor type response as a convenient vehicle for testing other hypothesis, usually involving procedural variables which have been derived frogeneral behavioral learning theories.

Trumbo points out that the rotary pursuit apparatus is a good case in point. He states that it is certainly the best known and most widely used motor skills apparatus available. It is, of course, a motor tracking task and it has been used in a host of studies on distribution of practice and other procedural variables. Yet, in turns of task variables it is limited to little more than rate of turn, target size, and stylus weight. The rotary pursuit generally yields a single time-on-target score. The time-on-target score has definite limitations in that it does not use all of the data in the error distribution. In order to make use of the data in the error distribution, it is necessary to measure root mean square error or average error or integrated error. These values are readily obtained with an electronic scoring system.

Bahrick, Fitts, and Briggs (5) point out in a 1957 paper that time on target scores, which are the amount of time during a particular trial that a person is able to remain within an arbitrarially specified region around a target using some kind of tracking device, presents some real problems in turns of the derivation of learning curves and their meaningfulness relative to the basic process of vigilance, tracking and monitoring. They advocate the use of the root mean square measurement (RMS) as the best method of avoiding difficulties because it simply substitutes a single function for an unlimited number of functions determined by all possible target or activity dimensions. In general, response characteristics may follow a continuous and normal distribution, but learning or practice results in a diminished variance of this distribution; however, performance is scored according to an all or none criterion of frequency of occurrence. This scoring practice accounts for the lack of predictability of such tests as the steadiness test, the dotting test, tweezer dexterity tests, pegboard tests, etc., whenever success is scored against an all or none criterion.

Kennedy (6) reports on a vigilance task which increased in difficulty from one to three channels. When he stressed a group performing both one and three channel vigilance monitoring, the stress group performed better than the nonstress indicating a certain level of arousal or an alerting response to the threat of shock.

In another study, Demaio, et. al., (7) reported comparisons between student and instructor pilots using a visual scanning task, showing that instructor pilots learned to attend to critical features more efficiently than do individuals with little or no flight experience. This suggests the interesting possibilities of using a variety of scanning tasks in the undergraduate pilot training program to facilitate the more rapid development of adaptive scanning strategies. Thus, instead of using a scanning or monitoring type task to evaluate stress or other such factors, we have the use of the task as a training device.

Reaction time has also suffered in the total context of suspicion about the ambiguity of simple performance measures. This again relates to the concept of performance versus effort. This can be seen in the form of simple performance in two working situations which are obviously different because the operator sees it as his task to achieve a particular output rate and he adjusts his effort accordingly. Now this effort can be detected by various refinements of performance data, but it remains true that straight forward speed-error data will not easily reveal difference of effort. Using speed as a measure of complexity makes unwarranted assumptions about the essentially sequential rather than parallel nature of human performance processing. Nevertheless, reaction time measurement has been an attractive area for scientific research for some time. Apparently scientists have been intrigued with the attempt to quantify the absolute speed to which a human can react to a signal. Perhaps the landmark paper in this field was one by F. C. Donders in 1865. O'Donnell (8) reports that Donders developed a "subtraction method" of reaction time. Basically, this theory stated that the elements of the reaction time response were addictive, in that each component of the response begins immediately when the only, when the preceeding component has ended. Thus, decision "is a component of a reaction involving choice. When the decision is being made, nothing else is going on and when that process ends, the subjective immediately moves into another, perhaps, movement phase, if this and other assumptions are true than decision time could be calculated by knowing the total reaction time and substracting from it the time required for all other components of the response." Using this approach, Donders distinguished between three types of reactions. The "A" reaction involved a single response to a single invariant stimulus. This is what is now called simple reaction time. The "B" reaction involves two stimuli and two responses with the stimulus response relationship always constant. This is the most simple form of choice reaction time presented. The "C" reaction also used two stimuli but in this case only one response was used and that response was to be given to one of the stimuli, but not to other. Consequently, one could calculate, response selection time by calculating "B" minus "C" and stimulus categorization time by "C" minus "A", etc. In spite of over 100 years of use of reaction time testing with various controversies we still have many measures that are only indirectly related to reaction time per se, but they use reaction time as a measure of performance. These secondary techniques, if we may call them that, in general differ from the classic few of reaction time where the stimulus is a relative discrete signal introduced by the experimenter for the subject's discrete response. These secondary techniques involve self-pacing by the subjects, simultaneous performance on a number of tasks and/or use of a complete series of reactions to obtain a total score for task completion rather than a reaction time score per se. For example, Hartman and I have used reaction time component for mental arithmetic and tracking, etc. Recently, Wood (9) has undertaken a study of the neurophysiological basis of reaction time change as a viable means of exploring physiological mechanisms of local muscular fatigue and fatigue effects on sensori-motor performance. Here, he used measures of reaction time, evoked potential, and EMG as indicators of central and peripheral activity. With this particular reaction time model he is able to fractionate total reaction time into component latencies, he is able to study central versus peripheral issues which are featured predominantly in both fatigue and reaction time. Wilkinson (10) reports on a small battery-powered fully portable device for administering a four choice serial reaction time test and recording the results on a standard magnetic tape cassette. In preliminary performance trials, this test appears to reflect fatigue due to continuous repetitive responding in a way similar to classical nonportable multiple choice serial reaction tests. Gaillard, et. al., (11) reports some effects of ACTH 4-10 using a serial reaction task, concluding that this particular drug counteracts the usual decay and performance as a function of time on task due to increasing boredom and mental fatigue. Bartz (12) describes an experiment using peripheral detection and central task complexity where reaction time is measured relative to the peripheral stimulus. He supports Hebb's arousal theory which would predict that increasing the complexity of the central task would heighten the subjects' vigilance performance. Salzman and Jaques (13) explored the relationship between heart rate changes and reaction time. They found no relationships between reaction time and the heart beat immediately preceding the stimulus or with the beat during which the stimulus was presented. Therefore, response latencies in terms of reaction time did not differ significantly as a function of phase in the cardiac cycle as predicted by J. Lacey who suggested that feedback from cardiac events can effect central functioning by a negative feedback regulatory loop mechanism. Thackray, Bailey, and Touchstone (14) reporting on boredom and monotony while performing a simulated radar control task showed that a high boredom/motonony group revealed greater increases in response times, heart rate variability, and "strain", and a greater decrease in attentiveness. They conclude that the pattern associated with boredom and monotony seems more closley related to attentional processes than to arousal. Holt and Brainard (15) reported an experiment using reaction time and a condition of selective hyperthermia where they raised cortical temperatures. In tis task, (a simple choice reaction time task) response times and response variabilities were decreased compared to performance in either control or placebo condition.

Physiological Correlates: Although it has always been recognized that both psychologists and physiologists [illegible] skills, there has been a tendency to think of some problems of energy expenditure, heart rate, etc., as physilogical and others such as we have just discussed, vigilance, tracking, monitoring, as psychological. However, as Singleton (16) has pointed out, it now begins to looks as though the complexities and interactions within the human body are such that neither discipline is adequate alone for the study of any problem of man at work. The physiologist can be accused of too narrow an approach with insufficient regard to cortical dominance and tending to deal with endocrine and autonomic parameters. Similarly, the psychologist can be accused of treating the human operator as too "pure" an information processing device without sufficient regard for subcortical and somatic factors which clearly influence performance. Directing attention now to the physiological correlates we find that physiological measures of heart rate, muscle physiolcgy, body metabolites such as 17 keto-steroids, etc., have been used to provide a method of measurement and to provide a set of standards. In terms of a set of standards derived from these kinds of measures it must be pointed out that it has been difficult to determine that a partciular task requires an energy expenditure of so many calories per minute or hour, but it is even more difficult to determine whether this energy expenditure, or heart rate level, or outpouring of metabolite constitutes a light workload, a light energy stress, or a heavy, or even an intolerable amount of effort on the part of the human. Another aspect of the general difficulty about physiological measures is that the stress on the operator tends to have similar effects whether it is due to work, development, fear or environmental factors such as noise, vibration, etc. Nevertheless, there is some value in using physiological concepts to attempt to predict the behavior of the human operator keeping in mind that the relationship between simple physical measures of the environment and the corresponding effects upon the operator invariably turns out to be a multidimensional problem with dominant influences from many variables difficult to measure or control especially those in the psychologic realm of attitude and motivation. Sharkey, McDonald and Corbridge (17) point out in a paper in which they evaluate pulse rate and pulmonary ventilation as predictors of human energy cost that this human energy cost and efficiency are of considerable importance in the evaluation of the equipment for industrial tasks. Now this is particularly true if we look at industrial tasks and related equipment in the light of aircrew protection garments for altitude effects, thermal effects, and chemical defense. Pulmonary ventilation rate as a predictor of human enery costs has long been known and used; however, the accurate assessment of ventilation rate depends on cumbersome gas analysis techniques and still requires that gas be collected in the field and transported to the laboratory where the time spent in analysis still restricts sample sizes to those relatively small. Therefore, this particualr investigation attempts to compare the precision of prediction of human energy costs afforded by both pulse rate and ventilation rates. In short, in spite of the attractiveness of using relatively simple determination and recording of pulse rate, the use of pulse rate alone in lieu of ventilation rate would indicate the possibility of larger errors in predicting energy cost. In spite of the drawbacks of pulse rate alone, it should be pointed out that Sharkey, <u>et. al.</u>, indicate that predicted energy costs were over-estimated rather than under-estimated (17).

In another study relating to task and load difficulties using the EKG by Schwarz and Ekkers (18), they discuss the task of developing the optimal functioning reliability of a complex system from three aspects (1) the development of analysis of the reliability of the system (2) the organizational rules of procedure by which unanticipated emergencies can be forestalled and (3) equipping the individual operator physically and mentally to regulate tasks and load difficulties. In this study, they found that EKG was significantly related to the perceived gravity of an unannounced or emergency situation. In another study relating task demand reflected in physiological variables, Frakenhaeuser and Johansson (19) measured catecholmine excretion and heart rate variance pointing out that the physiological arousal indices were more susceptible than performance measures to the level of task demands. In other words, the higher demand imposed by a double conflict task was reflected in relatively larger increases of adrenalin excretion and heart rate where as performance measures which were psychological remained unaffected. A study by George Montomgery (20) on the effects of performance evaluation and anxiety on cardiac response in anticipation of a difficult problem solving task showed that analysis of second-by-second changes in cardiac rate revealed that waveform components were sensitive to both anxiety and failure within the evaluation stress condition only. Initial cardiac acceleration responses covaried with performance measures across anxiety groups apparently reflecting differences in confidence or motivation. Concept of the anticipated problem solving task was reflected in a cardiac foreperiod deceleration response which is very likely related to attentional readiness for the beginning of the problem.

In a relatively long-term study of the activity of the nervous system during pilotage activities of letdown, approach, and landing Nicholson and his colleagues (21) have related pilot subjective assessment of his workload to changes in heart rate using the RR interval and the finger tremor measured by an accelerameter. They have concluded that the mean heart rate interval around touchdown reflects the workload of the crew's letdown, approach, and landing phases whereas changes in finger tremor are associated with untoward events during the approach which relate to difficulties in the dynamic flight situation involving weather, wind shear and other factors. A follow-on study of four years of workload assessment was done to determine how effective their measures were in terms of reliability. They report that the subjective assessments of the pilot are meaningful. However, they note that the degree of neurological changes associated with the possibility of impaired subjective analysis of workload may be related to the fact that under difficult circumstances a pilot may have a degree of central nervous system arousal above that which may be associated with optimum performance. Their finger tremor technique is interesting because it may be related to the release of catecholmines which are in turn associated with finger tremor. On the other hand, ballisto-cardiographic effects may play a role in this mechanism, since finger tremor has been observed with muscular contraction during pronounced tachycardia. As they conclude, "the peripheral changes in nervous activity observed during the letdown, approach, and landing may indicate two physiological states both of which arise from central nervous arousal. In the case of high workload letdowns without untoward events, profound cardiac acceleration and limited finger tremor are the physiological changes of neurogenic origin. In letdowns in which the approach is complicated, profound finger tremor dominates the picture and may be associated with circulating catecholmines."

In another study relating physiological correlates to changes during a mental task, Kahneman, Tursky, Shairo and Crider (22) had subjects perform a paced mental task at three levels of difficulty while they recorded pupil diameter, heart-rate, and skin resistance changes. They reported a similar pattern of sympathetic-like increase found in the three autonomic functions during performance intake and processing followed by decrease during the report phase. The peak response of each of these three measures was ordered as a function of task difficulty. There is considerable evidence that problem solving performance as well as other tasks are associated with activation of the sympathetic nervous system indicated by increased electro-dermal activity, increased heart rate, increased blood pressure and peripheral vasoconstriction.

It has been long known that pupil dilation occurs during mental activity. More recent research by Kahneman and Beatty (23) suggests that this indicator may be particularily sensitive to mental activity in a special way. While it is true that pupillary changes are associated with activitation of the sympathetic nervous system an even more important index of arousal is the fact that the oculomotor nerves which act to change pupillary size originate in the ascending reticular formation and provide us with an important window into that system. It is unfortunate indeed that pupillary measures are so difficult to obtain in terms of equipment and the physical constraints imposed. Nevertheless, important work is underway that will hopefully relate pupillary changes to other more easily obtainable physiological and psychophysiological correlates of workload, performance, and stress. Before turning to these phychophysiological correlates, we will direct immediate attention and comment to some of the remaining physiological correlates, namely body metabolites, and the electromyograph.

By measuring simultaneously the urinary excretion of most of the known hormones, it has been established that the organism's response to stress involves a total neuroendocrine apparatus. As Dukes-Dobos (24) has stated, these hormones can be divided into groups according to their excretory pattern. One group of hormones is excreted in increased amounts during the stress exposure and the other group shows a biphasic change inasmuch as these hormones are excreted in decreased amounts during the stress and in increased amounts during the recovery phase. Studies performed on the urinary mucoproteins suggest that the excretion rate of this substance is an indicator of the speed of catabolic processes in the body reflecting the balance of the total neuroendrocrine response to stress. These measures while important present us with certain problems in the interruption of such changes. For instance, we only know that the individual has been stressed; we do not know the exact time in which the stress occurred. Also, a reduction in excretion after repeated exposures to a stress may be due to either adaptation or to fatigue.

Since the active state of the human operator is connected with the sympathetic tonus, one could assume that the hormones of the sympatho-adrenal-medullary system (adrenalin and noradrenalin) must always be excreted in increased amounts during physical exercise or mental work. However, as Dukes-Dobos points out, while some investigators have found increases in one or the other cathecolmines in the urine after physical exercise or physchological stress others did not find such changes at all. One reason for the confusing results may be that the blood-brain barier permits only a small amount of noradrenalin to cross through from the brain to the blood and then show up in the urine. Therefore, the urinary noradrenalin level depends upon the activity of the peripheral sympathetic nerve endings which may or may not be related to noradrenalin release in the brain. Thus, urinary noradrenalin does not give a reliable estimate of the total noradrenalin excreted in the sympathetic nervous system. On the other hand, urinary excretion of adrenalin may reflect completely the activity of the adrenal medulla. According to the classic experiments of Von Euler (25) excretion of catecholmines will increase after physical exercise only if the subject discovers that the performance requires a special effort.

In one of the many studies on airplane pilots performed by Hale (26) urine was sampled over a 28-hour period every four hours from the crew members during the first transatlantic helicopter flight. The flight was a risky undertaking and bad weather conditions often threatened its success. The average adrenalin and noradrenalin excretions of the crew members were elevated. What can be considered a unique finding was that an increased adrenalin excretion during the flight was observed in all ten of the subjects. Other urinary metabolites measured were excreted in increased amounts by some subjects and in decreased amounts by others compared to the controls. Thus, adrenalin excretion seems to be the best parameter for accessing the magnitude of stress brought about by a task which is not demanding as far as physical exertion is concerned, but is connected with stressful work conditions and is in fact hazardous.

Reflecting Selye's (27) stress concept, physical as well as mental work can be considered as a stress factor which may evoke the general adaptation syndrome, thus, activating the pituitary-adrenal-cortical axis. Many studies have demonstrated that the excreted metabolites of this endocrine system show quantitative changes after physical exercise as well as psychological stress. This mechanism has been explored by studies utilizing measurements of 17-hydroxycorticosteroids (17-OHCS). In general, we have found that during periods of stress, the 17-OHCS excretion increases; however, we have also found that an exposure of day-to-day stress may bring about a state of "chronic-adaptation" fatigue which will cause a drop in 17-OCHS excretion instead of an elevation. A relatively unique tool for the evaluation of 17-OCHS levels was pioneered and developed by Shannon (28) using parotid fluid collection and analyzing these samples for free 17-OCHS levels. The development of this technique is very interesting to follow and represents a determined effort to avoid some of the dangers, discomforts, and logistics problems involved in collecting in-flight specimens of blood and urine. The refined technique for collecting parotid fluid involves a plastic collecting device using an acrylic bite-block molded to the individual bite of each subject. This technique allows easy and rapid self-positioning of the device over the parotid duct opening. It should be noted that in a study by Warren, Ware, Shannon and Leverett (29) they state that the in-flight parotid fluid collection technique has been developed to the point where it represents a valuable adjunct for in-flight physiological studies. This is especially true because rises in steroid levels in parotid fluid does not demonstrate the lag that is characteristic of urinary steroid responses.

The electrical measurement of muscle activity, the electromyograph, has a long history. Most of the derived muscle physiology relates to laboratory studies in which the particular muscle of muscle group has been stressed to bear maximum muscle contractions during a relative short period of time. Not many muscle studies have been related to operational purposes where muscles are stretched, fatigued or otherwise tasked over long periods of time and unique workload situations. If we set aside the work of those various researchers and clinicians interested in muscular skeletal relaxation techniques perhaps the most important investigations in the neurophysiology of muscle has been done by Basmajian (30). We will not review all of his work because most of it is of a basic nature and is not of immediate operational interest. However, control that can be obtained by humans to indicate that this important physiological correlate should not remain ignored in the assessment of workload, performance, and stress.

Basmajian has developed a technique and the necessary instrumentation in the form of bipolar intramuscular electrodes to study the changing patterns of activity of individual motoneurons through the application of modified electromyographic methods. In a series of studies he has demonstrated that human beings can learn to activate or repress any number of spinal motoneurons in a given pool. The human subject can also learn to voluntarily select individual motoneurons and to control the firing of these neurons through the assistance of auditory and visual feedback. Some of his subjects learned such exquisite control over these individual motoneurons that they were able to produce various rhythms and patterns by deliberately speeding and slowing the firing of the individual neuron. Basmajian's technique is to be a promising method for the study of many fundamental phenomona in the nervous system relating to cortical and subcortical effects upon the motoneuron and the of conditioning and learning. One other important area of investigation would be that of pharmacological agents on various parts of the motor pathways and muscle activity itself. Of a more applied nature in the area of electromyographic investigation, Lafevers (31) reports on a work task performed in a full pressure suit. Lafevers performed a power spectral density analysis of EMG recordings from several muscle groups involved in a push-pull task at various reach positions in both a space suit and in shirt sleeves. He feels that the power spectral shifts indicated significant findings relative to the performance and stress requirements for these muscle groups. The reason that these would not be expected to appear in ordinary electromyographic determinations is the fact that the task requirements of the study were not of a fatiguing nature nor did they stress the muscle groups to their upmost. This type of work suggests many areas where the relationship between the man-machine interface in terms of motor activity and response requirements might be explored. Here, it should be possible to identify task requirements that promote muscular fatigue and the resultant effects of this fatigue on both man and the particular task involved.

Stress Correlates: The stress correlates of workload and performance might be considered the environmental, operational, and internal results of acute, chronic and cumulative effects of psychological and physiological activities. In response to environmental stressors such as heat, noise, vibration and so fosth, and the operational requirements of a particular task, duty, or mission the internal environment of the human begins to respond in a more or less predictable fashion. The end result of the stress correlates of behavior is a deterioration in activity and a series of determinable changes which we usually call fatigue. Fatigue, as Grandjean and Kogi (32) report in their introductory remarks to the Kyoto symposium on methology of fatigue assessment, is a subjective sensation in many ways where we feel not only tired in our bodily parts and clumsy in psychomotor activity, but we feel hampered and inhibited in doing either physical or mental work. This inhibition of activity continues until we are constrained against doing any form of active endeavor. These sensations of fatigue can be assumed to have a protective function in that they force us to avoid further stress and allow recovery to take place. The concept of fatigue is not a popular scientific term because it is difficult to evaluate and to quantify. A series of studies reported by Wolf (33) and later by Saito (34) report that the sensation of fatigue has three major components: (1) a sensation of bodily tiredness and drowiness; (2) a sensation of weakened motivation or concentration towards a task and, (3) a group of physical complaints that relate very closely to what are commonly called the psychosomatic disorders. These psychosomatic complaints are usually those of headache, palpitations, tachycardia, shortage of breath, loss of appetite and indigestion or sleeplessness. A predominance of these kinds of complaints is usually referred to as clinical fatigue. In the presence of clinical fatigue, absences from work predominate due to "illenss," and there arises a general negative attitude towards one's work, one's superiors or the place of work which obviously can just as well be a cause of clinical fatigue as well as be a result of it. Compounding the problem of evaluating chronic fatigue is the fact that clinically it is well known that people with psychological conflicts and difficulties are especially prone to this state. This makes it difficult to separate the psychogenic factors from exogenous causes of fatigue. In spite of the fact that the actual components of fatigue are somewhat difficult to scientically quantify, it is not difficult to assume that the commonly experienced sensations of fatigue are very likely a biological sign of the necessity for man to enter into a recovery phase by informing us that the relative inflow of fatigue is exceeding our capacity. As Grandjean and Kogi report the following signs are observed in conditons of chronic fatigue: (1) a general weakness and drive and loss of initiative; (2) a tendency to depression associated with unmotivated worries; (3) increased irritability and intolerance (occasionally exhibited with unsociable behaviors).

In considering the stress correlates of workload and fatigue we need tobe aware of the role of the activating system and inhibiting systems of the CNS. We know that the brain contains neural structures responsible for maintaining wakefulness and alerting the cortex. It has been shown that lesions of the medial mid-brain make animals inattentive with low motivation and drowiness. This structure is located in the reticular formation of the mid-brain and is called the activating system. Stimulation of this system arouses the individual or animal, while destruction of it causes the animal to go into a permanent coma. There are also neural pathways leading impulses from the ceberal cortex back to the activating system. These corticofugal pathways converging on the reticular formation have a function similar to a feedback system, that is, impulses originating in the cortex are capable, through this feedback, of stimulating the ascending reticular activating system which in turn maintains the cortex and the behavior of the organism in a state of arousal and alertness. All of the classical afferent pathways coming from the sensory organs send collateral impulses to the reticular activating system. This means that impulses from the environment, through the sense organs, or from muscle activities, can stimulate the ascending activating system and thereby increase cortical activity. Though there is recent evidence that lateral brain stem regions are as important as medial regions for attention and arousal, it is generally admitted that unspecific neurons decisively regulate arousal and attention. Related to the neurological aspects of workload/fatigue are other investigations that have shown that stimulation of the activating system can spread to the autonomic nervous system giving rise to hormonal changes in the internal organs such that the organism may poise itself for energy expenditure.

The work of Hess (35) showed that electrical stimulation through chronically implanted electrodes produced a tendency to fall asleep and to produce pronounced muscular relaxation in cats. This discovery later confirmed by many others has shown a active inhibition mechanism which spreads from the subcortical structures to the ceberal cortex and acts to depress cortical functions. These systems have a direct depressing influence on the ascending reticular activating systems. Therefore, cortical inhibition can result from two different causes. On the one hand, cortical activity may decrease as a result of lowered sensory inputs or a lowered corticofugual feedback. This might be called a passive inhibition. On the other hand, cortical activity can be reduced by an active inhibitory function which would be elicited by increased activity of the inhibitory system. It is interesting to note that we see changes in the brain wave which involve a flattening of electrical activity which are associated with suppressed behavior in both fatigue, in states of chronic anxiety, and in certain drug effects which act to suppress the central nervous system. It is important also to remember that the organism regulates its feelings of fatigue or relative arousal not only through the neuralmechanisms, but through endocrine factors which are ultimately responsible for maintaining a certain functional state for hours or longer periods of time.

In spite of the neurological and endocrinologic relationships discovered and understood, it is still a problem for us to remember that fatigue is still subjectively evaluated. In other words, just because the conditions of fatigue exist does not necessarily mean that performance decrement occurs. It may be true that subjective feelings will preceed a loss in performance ability, but not necessarily. It is well known that in spite of great fatigue the human organism will response with adequate, if not, lifesaving performance levels. Further, we know that there is a situation of unacceptable fatigue which people classify as a kind of fatigue called overwork, overload, or exhaustion or other kinds of terms. Since these kinds of fatigue concepts are related to subjective judgment we, therefore, get back to the psychophysiological implications of workload, performance, and stress. We will see that some investigators in assessing fatigue feel that there are indexes of fatigue through physiological measures such as increase in heart rate, reduction of sinus arrythmia, and so forth. Regardless of whether one is concerned with the physical aspects of fatigue or the mental aspects of fatigue, certain symptoms can be considered as a consequence of cortical inhibition activity. The following symptoms of what might be called "cortical fatigue" are those which need to be evaluated, (1) decrease of attention, (2) slow and impaired perception, (3) impairment of thinking, (4) decreased motivation, (5) decreased performance speed, (6) decreased accuracy, and (7) decreased performance reserve for physical and mental activity. While most of these factors have been investigated to some extent it is probably the electrical activity of the cortex which may give a better picture of activity which can be considered as having a direct regulatory effect.

In a factor analytic study of mental fatigue, Kogi and Saito (36) were able to demonstrate that certain changes in cortical functions were related to various phases of a 24-hour period. The measure of cortical activity they selected was the critical flicker fusion test, but changes in CFF were also reinforced by changes in a choice reaction time test.

A study by Ettema and Zielhuis (37) investigating the physiological parameters of mental load demonstrated that a simple, binary choice test providing different mental loads or levels of difficulty, showed systematic changes in heart rate, sinsus arrhythmia, systolic and diastolic blood pressure, and rate of respiration.

Kashiwagi (38) was able to constuct a fatigue rating scale which allows a judgment of human fatigue through a person's appearance. The use of such a scale might be very helpful in the field as far as management or field commanders are concerned and might be of use to some of the mission crew fatigue studies done at the School of Aerospace Medicine by William Storm and his colleagues. Presently, subjective fatigue and sleep data are collected from various mission groups. These measures are used to assess the overall effects of mission requirements upon sleep loss and workload requirements (39).

A system using a concept of task-induced stress was developed by this author and used in the stress testing of special mission personnel in the U.S. Air Force (40). This concept was structured around the tasks an operator must perform in an advanced space system. He must perform a relatively large array of discrete, discontinuous operations against a background of monitoring and information processing tasks. In terms of information theory, the discrete, discontinuous functions would constitute a source of noise in the form of unwanted or distracting signals when the operator was trying to monitor and process a continuous input. Increasing the signal rate of the discontinuous tasks makes the detection and identification of the continuous task more difficult in the same manner that increased noise acts to degrade audible signal detection and recognition. By structuring the task situation so that the operator is uncertain as to what is signal and what is noise, it is possible to cause him to continually shift his attention from signal to noise and noise to signal. This is the nature of competing tasks and the end result of such a situation can be regarded as task-induced stress. By further structuring the situation so that the operator has been allowed to find out that he can in fact perform both the discrete, discontinuous task and the continuous monitoring task independently, he is quite apt to assume that he should be able to do them together with perhaps only a little more effort. When he finds out that he has much difficulty doing both tasks he is led to conclude that there is something wrong with <u>him</u> and he would much better if he could only find the optimal technique or "a gimmick". This sort of structuring tends to invite the formation of internalized, psychologic stress which is not relieved much by hostility towards the tasks themselves. Since there is no obvious source of the proficiency problem presented by the competing task, the psychologic feelings generated by failure to perform well tend to be self-directed rather than task directed. In this situation, the induced stress is more than the sum of the stress of performing each task independently. The results indicated that a criterion group of those finally selected for the special mission using various other criteria was better able to adapt to the two competing tasks and was less susceptible to the signal noise, ambiguity and the induced task stress than the special mission personnel group as a whole.

It must be noted that the evaluation of highly specialized groups selected by virtue of years of experience and special talents presents a unique problem to those of us charged with the responsibilites of investigating the effects workload, performance, and stress within and upon the human operator.

Another important area of performance decrement in relationship to environmental and operational working conditions is that of exposure to hot and humid environments. In an important study reported by Welch, Longley and Lomaev (41) they found that sweat loss and pulse rate were found to be unreliable methods of measuring fatigue and that skin temperature was completely unreliable as an index of fatigue except when the temperature and humidity were high. Their experiments confirmed that a rectal temperature of 38.8° C to 38.9° C will in most cases coincide closely with the onset of actual exhaustion. Another study in this area by Grivel (42) indicates that the permanent, specific heat effects on psychomotor and mental performance are related to preferentially in that different aspects of the same activity were considered to determine the effects of climatic stress. In other words, heat acts differently on the reactivity aspects of performance than those aspects of performance associated with continuous attention. In the field of time-varying heat effects, studies have examined the possible transitory effects of heat as well as long term evolution of effects found at the time of first heat exposure. Here, a sequence of events can be distinguished, each characterized by a particular kind of ambient heat effect upon performance. This suggests some type of learning or conditioning takes place relevant to heat stress.

The Psychophysiological Correlates: We have seen so far that we have both common and scientific knowledge that individual and combined streses, both physiological and psychological, can adversely effect mental performance and judgement as well as physical performance. We have just discussed how both hyperthermia and fatigue can produce deteriorated, objective judgements regarding environmental situations as well as degraded performance. The individual's subjective judgement or insight concerning the quality of his own performance is similarly degraded. With sustained exposure to stress, he tends to overestimate his capability and to discount his errors. Subjective identification of degraded central nervous system (CNS) function is generally based not on recognition of degraded performance, but on secondary indicators such as dimming of vision in case of hypoxia and reduced span of attention in the case of fatigue just to use one example. We are primarily interested in looking at psychophysiological parameters which relate to central nervous system function. Our interest stems from the fact that to the extent that these CNS changes are detectable through analysis of peripheral physiologic measures they can provide a sort of warning system of primary higher CNS functional decrement in the same sense that an oxygen partial pressure meter provides primary hypoxia warning.

Many years ago the Cambridge cockpit series of performance/fatigue studies examined behavioral changes observable through several hours of continuous performance when subject's "flew" a specially instrumented simulated aircraft. Bartlett (43) summarized these studies in terms of skill-fatigue effects. The experiments showed beyond question that, under the conditions imposed, "operator fatigue" does occur though in most cases the operator himself did not realize it. Within a maximum of 8 hours of simulator operating, the experimenter concluded that the subjects were still able to perform the operations, but only if they were especially careful to avoid known deficiencies characteristic of fatigue. Some inexperienced subjects developed significant deterioration of performance after only 1½ hours. One highly motivated, experienced subject went 8 hours without appreciable deterioration. Most subjects, as fatigue progressed, showed a lack of coordination between the recognition of the required operation and the necessary response. This is related most logically to impairment of the integrative function of the association areas of the cerebral cortex. Marked increases in lability and irritability were also observed along with changes in judgement. As time progressed during the performance, the number of small errors increased, but were later replaced by large errors. This was interpreted as reflecting degraded neuromuscular control with increasing levels of frustration. This was compensated for by a judgemental change in the standards of accuracy. The subjects were unaware of this change in their judgement of acceptable performance unless it was called to their attention. This idea of subjective lack of awareness is crucial in the operation of high performance man-machine systems.

Other studies of fatigue have demonstrated that subjects become tired of a specific task and show rejuvenated performance upon changing tasks. From a neurophysiological standpoint, this relates to a reduced level of general CNS activity upon habituation to a monotenous task. However, with the introduction of a novel stimulus there is a marked increase in CNS activity. This is the so called arousal or activation response. Subjective appraisal of performance and its relation to objective criteria under conditions of fatigue produced by prolonged wakefulness using skin resistance measurements as well as EEG tracing has been reported by Burch and Greiner (44). Generally, they found the subjective evaluations showed a high correlation with their bioelectrical measurements during pre-experimental control periods and the earlier portions of fatigue. However, as the fatigue progressed the subject's ability to evaluate his own state of consciousness begins to break down.

Several studies of mild hypoxia at the School of Aerospace Medicine have shown frequent lapses in simple performance tasks lasting only a few seconds and suggestive of a momentary loss of awareness. In fact, one of the tasks incorporated into a multielement psychomotor test device previously referred to as Neptune (an acronym for "neuropsychiatric test unit") was designed to provide a relative measure of operator consciousness during periods of experimentally induced hypoxia. This task was called Auditory Monitor and involved monitoring three Morse Code signals "A", "N," and "M" which are played in random order at a preselected speed. This task indicated momentarily loss of awareness and these periods of loss correlated with EEG changes indicating reduced cortical arousal.

The stress of sleep deprivation also shows brief, often dramatic intermittent pauses or lapses in ongoing behavior. Many studies show that these lapses increase in frequency, duration, and depth as sleep loss increases. Between lapses subjects are able to think and act under challenge almost as well as under preexperimental conditions. As lapses deepen, it is increasingly difficult for the subject to hold a stable frame of reference while performing a series of mental operations. If a deep spell of drowsiness occurs in the middle of a serial operation, the subject will stop the sequence for a brief time and often lose track of where he had been in the series. Luby (45) has attributed decrement 14 psychological test performance under conditions of 118 to 120 hours of sleeplessness to fluctuations of attention. The frequency of periods of inattention increases as a function of the hours of sleep loss.

In comparing measures of physiologic activity with observed behavior it is necessary to note that an organized psychomotor response pattern involves three factors which must be integrated at a relative high cortical level. These factors are: (1) detection of the signal, (2) selection of the response, and (3) execution. Under conditions of disorganization the response pattern is fractionated rather than coordinated so that certain indications of response pathology occur. For example, under stress conditions detection may be accompanied by a startle response of varying degree to signals having high attention value. Signal detection may be degraded by a breakdown of scanning behavior as the attention span is attenuated. The wrong response may be selected and execution may be characterized by gross spatial errors and psychomotor movement, that is, moving first to the general area of the control and then to the control itself. Other execution errors involve operating the wrong control or using the proper control incorrectly.

In summary then, we see that various physiological and psychologic stressors individually produce variable degrees of decrement and behavioral performance, some of which are predictable. However, in combination, the effects of these stressors on performance become difficult, if not impossible to predict. Nevertheless, it is possible to monitor neurophysiological states and events to the extent that it is possible to identify CNS functional changes related to the primary cause of performance decrement. At the present state of the art, the likelihood of detecting a specific erroneous judgement using CNS functional criteria is not possible. However, it is possible at present to detect specific mental activity related to specific external and internal processing events. We will discuss these at a later point. At present we will deal with the identification of a CNS functional state correlated with unreliable or pathologic performance and judgement. It is interesting to speculate upon the fact that many of the early symptoms of some organic brain diseases and some focal brain disorders or not unlike some of the behavioral changes just mentioned and yet to date no orgauized study has been made relating psychophysiological variables with symptomatic or psychometric factors in such disorders.

Psychophysiological Monitoring of CNS Function

The physiologic evaluation of central nervous system function can be approached from two points of view: (1) the general state of arousal or level of consciousness, and (2) the quantitative and qualitative aspects of indivudal or specific CNS responses. We will discuss the first approach from a neurophysiological standpoint and relate it to some existing data under the title, "General Levels of CNS Activation". The other approach will be discussed under the heading, "Individual CNS Response."

Before delving into CNS monitoring itself, we should be aware of some of the considerations in terms of measures and analysis. Individual physiologic measures can be analyzed from at least two aspects: (1) averaged or integrated values, and (2) quantitative analysis which reveal changes due to individual CNS responses. These methods have been devised to permit a reduction of data to speed analysis and to allow for computer handling of the data. Four such measures, the electroencephalogram, the electrocardiogram, respiration and electrodermal responses will be discussed in some detail. In each case we will try to relate the two kinds of analysis to the two corresponding types or modes of CNS function.

The Electroencephalogram: The potentials observed from scalp electrodes measure a part of the electrical activity that underlies superficial cerebral cortex. The specific areas of cerebral cortex are identified with primary, sensorimotor activity and with the integrative function of the association areas adjacent to these sensory areas. The sensory association areas play a major role in deriving meaning from the impulses received in the primary sensory areas. The frontal area contributes to the integration of the sensory association areas permitting abstract and conceptualization. One can thus expect observable changes in EEG patterns relating to changes in activity involving these higher mental functions. Much of the problem in intrepreting EEG patterns is due to the highly complex wave forms developed. Most forms of analysis have been borrowed from engineering approaches to vibration stress which also presents multiple frequency wave forms. In engineering terms this is called frequency spectrum analysis in which the various frequencies are partialled or split-out for individual analysis for a specific time period. A modification of this approach developed by Burch (46) shows promise. The Burch method analyzes EEG wave forms in the time domain expressed as major and minor periods. The major period represents the dominant EEG frequency for a specified time interval and is defined by baseline crosses of the raw EEG. The minor period represents the superimposed waves between the baseline crosses. Major and minor periods are each summed and represented as a total count during a time interval or epoch such as 10 seconds. The Burch method involves an additional display referred to as spectral analysis. This divides the raw EEG spectrum for each given epoch into 10 frequencies bands with reference to both major and minor periods. This analysis [illegible] in a form similar to the Grey Walter frequency analysis system. The readout provides a value for [illegible] of the 10 major and minor period frequency bands during every epoch. The amplitude of this writeout indicates the total time in the preceding epoch during which the analyzer detected periods with values falling within the frequency limit assigned to that particular band. The model frequency band for either the major or minor period is that band in which the greatest accumulated time is scored during the epoch. A total of major and minor period counts represents a characterization or signature of the EEG frequency spectrum during the epoch selected. Total counts over epoch's of a few seconds reflect individual reflexes involving a major portion of the cortex. Changes in modality of the frequency spectrum during the corresponding epochs are related to the quantitative aspects of such reflexes.

In contrast to the previously described frequency-period analysis which is concerned with time domain only, Riehl (47) developed a method which relates both time and amplitude domains. He defined an activation response which he called Ua. This can be written in the form of a equation where Ua equals F (the dominant frequency) multiplied by the reciprocal of the average amplitude. In this equation the dominant frequency is defined for the major period count and the average amplitude is that which is obtained by full wave rectification and integration. In order to derive this function, an analog computer is employed to integrate the wave form, and to obtain a continuous real-time representation. The integration of Ua over epochs of 10 seconds recorded at a relatively slow chart speed provides a convenient readout of an activation response over specific time periods. The Ua itself will exhibit major fluctuations of only a few seconds duration. These can be evaluated as identifiable CNS responses to known stimuli.

A more recent approach is to use power spectrum analysis using a fast Fourier transform. This yields a representation of extremely small power shifts over very small epochs. This result can be obtained on-line in terms of percentage of power in each selected frequency band width or in terms of pure power.

The Electrocardiogram: Nervous control of heart rate is classically described as mediated through vagal parasympathetic cardic-inhibitor fibers and through sympathetic cardioaccelerator fibers. The vagus nerve cardio-inhibitor fibers originate in the bilaterally paired dorsal motor nuclei of the Vagus. These nuclei lie in the floor of the fourth ventricle throughout most of the length of the medulla oblongata. The sympathetic cardioaccelerator fibers originate centrally in paired nuclei contained within the reticular substance of the Medulla. Beat by beat values of heart rate are obtained by measuring the period (R-R interval) of each cardiac cycle. An analysis of heart rate or trend, or accelerator, vs decelerator information may be obtained by averaging the frequency of a number of cardiac cycles. This is most conveniently done by using a cardiotachometer. The beat by beat analysis of cardiac rate represents a very promising method of observing individual CNS reflex responses. This analysis shows two contrasting patterns: (1) during sleep, the record consists almost exclusively of a rhythmic increase and decrease of heart rate coincident with respiration. This is referred to as respiratory coupling. (2) During periods of wakeful sensory motor activity, such as speaking, walking, etc., the beat by beat pattern of heart rate shows a preponderance of nonrespiratory coupling or decoupling showing frequent cardioaccelerator reflexes as opposed to those seen only occasionally during sleep. The number of premature ventricular contractions per unit time is observed to increase under conditions of stress. Other specific electocardiovascular changes have been reported in the literature, but at this time is not yet clear whether these are a function of direct nervous control or indirect humorial influences.

Respiration: Control of respiration is mediated through autonomic and voluntary pathways. The primary respiratory centers lie in the medulla oblongata and in the pons. The medullary centers are described as paired bilateral half-centers which include both an inspiratory and expiratory half-center on each side. The half-centers are contained within the medullary reticular substance. The pontine reticular formation contains an inhibitory pneunotaxic center and apneuistic center which exerts a strong tonic effect on the bulbar inspiratory center. Voluntary control of respiration originates in the cerebral cortex and is mediated through the hypothalamus. Both inhibitory and acceleratory cortical influences appear most specifically localized in the frontal cortex.

It is interesting to note that the medullary centers for respiratory control and cardioaccelerator control lie close to each other within the medullary reticular formation. Thus, it is not surprising that there should be a strong interaction between respiratory activity and heart rate. As we have noted during quiet periods of CNS activity heart rate is predominantly coupled to the respiratory cycle while during periods of CNS arousal the respiratory coupling is frequently replaced or decoupled by cadio-accelerator reflexes associated with brief respiratory arrest. Similar transient increases in heart rate are occasionally associated with a marked increase in respiratory rate. This observation suggests that the cardioaccelerator reflexes may derive from two clearly distinguishable neurophysiological mechanisms.

In a comprehensive review of sinus arrhythmia reflex mechanisms, Heymans cites clear evidence that in lower animals the cardiac vagal center is subject to two inhibitory, cardioaccelerator influences, one arising from the lungs exhibits increasing activity with mild pulmonary inflation and the other is mediated directly from the respiratory center (48). The latter influence is fully capable of producing typical sinus arrhythmia in the complete absence of pulmonary ventilation. This fact makes is reasonable to postulate changes in heart rate arising from cortical activity mediated directly through respiratory centers to the cardiac vagal center.

Electrodermal Responses: Electrodermal responses (EDR) which include galvanic skin response and the basal skin response, are predominantly, if not exclusively, mediated by the sympathetic nervous system which produces changes in skin resistance highly correlated with sweat gland activity, the so-called galvanic skin reflex.

A comprehensive review of galvanic skin reflex neurophysiology by Wang discusses stimulation, transsection, and oblation techniques employed in the CAT to identify CNS excitatory and inhibitory centers. The suprasegmental excitatory areas include the sensorimotor area of the cerebral cortex, the hypothalamus of the diencephalon, and the facilitatory reticular system in the diencephalon and mesencephalon. Two pathways of the GSR which are separate at the cortical and diencephalic level converge on the preganglionic sympathetic sudomotor neurons in the spinal cord. The facilitatory influence of the diencephalic and mesencephalic reticular system is characterized as follows: When the facilitatory reticular system is stimulated in both the interbrain and the midbrain, the response or effect varies with the strength of the stimulating current. Weak current elicits no response itself, but augments the reflex. Moderately strong currents evoke a small response itself and also enhances the reflex. A very strong current which calls forth a large response by itself, suppresses the reflex during and immediately after stimulation, but has a late, long lasting facilitatory effect on the reflex. This effect begins one minute after stimulation, reaches a peak in two or three minutes and then gradually declines to zero in 30 to 40 minutes.

The inhibitory centers identified include the frontal ceberal cortex, the caudate nucleus, the anterior cerebellar lobe and the bulbar medial reticular formation. The cerebral cortex has the least inhibitory effect and the bulbar medial reticular formation the strongest.

A number of stimuli characteristically elicit the GSR on selected sites of human skin. These include startle, painful or other strong sensory stimuli, violent respiratory activity, generalized muscular activity, and strong emotional stimuli. GSR activity during arousal conditions occur spontaneously in response to no apparent stimulus. This is the so-called nonspecific GSR. GSR's which occur in response to known stimuli are then called specific GSR's. The sensitivity to stimuli evidenced by the frequency and magnitude of GSR's are observed to fluctuate through relatively wide ranges in the course of normal daily activity. This suggests a threshold-type mechanism, that characteristically affects a wide span of control.

Individual GSRs are characterized by a transient drop in skin resistance in a period of a few seconds. This expression of reflex activity has been analysed in terms of the number of responses per unit time, of the amplitude of the individual responses, and response latency or the period of time between an administered stimulus and the onset of the GSR. Some time ago this author demonstrated that the area subtended by the recorded GSR, a measure which integrates both time and amplitude is a more sensitive indicator of stress than frequency or amplitude only (49).

Basal skin resistance or BSR varies slowly over a wide range as the individual fluctuates through states of consciousness on the sleep-arousal continuum. High resistance values are associated with low levels of consciousness such as sleep, and low resistance values with high levels as with intense excitement. Basal skin resistance has been observed to vary over a range of 10 to 1 in a period of 10 minutes during the period of transition from sleep to aroused wakefullness in the morning. The BSR tends to be lower during periods of frequent GSR and high during periods of infrequent GSR. The reduction of BSR during frequent GSR is apparently due to the cumulative drop in resistance resulting from the failure of complete recovery of the response mechanism to the prereflex level of resistance before the onset of the next stimulation. This same phenomenon is seem in the repeated stimulation of other neural response mechanisms. The relationship between BSR and GSR activity is apparently a function of the rate and magnitude of individual reflexes and the recovery rate of the skin resistance towards higher values. Thus, BSR can be seem as a form of integrated function of GSR activity. Analysis of both forms of electrodural activity provides further quantitative and qualitative information concerning some aspects of CNS reflex activity.

From this discussion, it is apparent that anatomically the central pathways mediating EDR provide numerous sources of influence upon the observed reflex. As a practical indicator of CNS function the enormous volume of GSR literature published emphasizes the fact that the GSR pattern produced is the result of multiple influences at the CNS level. To date these influences are rather poorly identified and their separate effects on GSR patterns are not clearly distinguished.

The discussion of these four measures referenced to CNS activity does not imply that other measures may not be of equal or greater value. Additional measures deserving consideration include blood pressure, pulse wave velocity, EMG (electromyography), eye motion (REM) as in dream studies, and pupillary measures. Further research efforts will be required to adequately determine the usefulness of the information given by each measure concerning the functional state of the central nervous system.

Central Nervous System Activation: With some insight into measurement procedures, we will look at general levels of CNS activation. The level of CNS activation or arousal relates to the state of consciousness normally ranging from deep sleep through wakefulness to intense arousal. Obviously, level of arousal is influenced by many factors including circadian periodicity, workload, emotional stimuli, and internal ideation. From a neurophysiologic standpoint, the state of consciousness is intimately related to the activity of the reticular formation. It, in turn, may be influenced strongly by afferent motor activity, the amount and kind of sensory stimulation, and the emotional state of the individual.

The Reticular Formation: Anatomically, the reticular formation occu es a central location in the brain stem joining the cerebral cortex with the spinal cord. It is composed of a network of interlaced fibers and contains nuclei surrounded as a group by the primary sensory and motor pathways connecting the cerebral cortex with the spinal cord. The central cephalic brain stem which includes the diencephalic and mesencephalic reticular formation is essential for awareness of the environment and voluntary purposeful movement.

In terms of the relationship of various CNS structures to conscious activity the ascending reticular activating system has been identified as having great functional signifance. Stimulation of the anterior portion of the reticular formation elicits electrocortical arousal in animals and has been used to produce wakefulness in human analeptics. This same anterior portion originates impulses distributed widely over most the cortical surface, particularly the association areas. It is interesting to note that there is a clear distinction between the mere meaning of impulses received in the primary sensory receptor areas of the cortex and the meaningful, purposeful activity evoked by concurrent activation of association areas, via the ascending reticular pathways. Thus, we find that impulses corresponding to a visual image arriving in the primary visual receptor area remain devoid of meaning unless the adjacent, visual assocation areas are concurrently activated. The arrival of sensory impulses devoid of meaningful association is characteristically demonstrated in sleep (or dreams).

The complexity of the relationship between the cortex and the reticular formation is emphasized by the important role played by the descending pathways which strongly influence the core of the brain stem. It is through these corticifugal pathways that emotional arousal and goal directed behavior of conscious processes are mediated.

The Limbic System: The functionally related neural structure called the limbic system surrounds the attachment of the cerebral hemispheres to the brain stem. This system is positively associated with the subjective and autonomic motor expression of emotion. Recordings of the electrical activity within the limbic structures have revealed two patterns of electrical discharge associated with excited behavior. This kind of behavior apparently involves the reinforcement mechanisms of the limbic system which serves both to increase the amplitude and to generalize the distribution of activity in other parts of the brain, including the reticular formation. Any information from physiologic measurements indicating the level of activation of the reticular formation should be helpful in determining the behavioral level of consciousness. We have previously noted the investigative window provided by pupilography into the relative state of the ascending reticular formation. Additonal physiologic patterns or identification of activity which would serve to indicate the contribution or involvement of the limbic sytem to arousal would help to determine its emotional component. At the present stage of development of biomedical monitoring most reports of physiologic measures obtained under stress present them in the form of average or integrated values over relatively long periods of time. The results generally correlate with trends in the level of CNS arousal. However, there is increasing evidence that detailed analysis of differences in the central processing of CNS responses is going to be evidenced by relative small transient changes in the EEG. These will be related to small changes in heart rate and electrodermal responses which together will provide more specific and reliable indicators of CNS arousal and the functional state of the central nervous system.

EEG Indicators: One of the most thoroughly studied features of the normal EEG is the 8 to 13 cycles per seconds alpha rhythm often observed most clearly over the occipital cortex. The complexity and multivariant nature of this rhythm has been demonstrated, to the extent that it is possible to distinguish individuals who demonstrate either persist alpha, responsive alpha or absence of alpha. However, it is necessary to realize that this is not a fixed classification and that individual alpha patterns actually encompass a continum in which "persistent" and "absent" types represent ends of the scale. It has been found that the sensory modality of the imagery characteristically employed by the individual largely determines his alpha rhythm. Visual imagery is associated with the absence of alpha, while non-visual, or auditory and tactile imagery is associated with persistent alpha and responsive or fluctuating alpha patterns are related to variance in the individual's imagery modality. This highly variable expression of alpha rhythm is further complicated by the fact that what appears as simple rhythm on a primary trace is really often a complex of frequencies from multiple sources. Finally, it is known that alpha may be absent due to nonspecific stress effects as seen in chronic neurotic anxiety states.

When present alpha rhythm appears most predominately in the relaxed, eyes closed, awake condition. However, it is possible to train a human being to produce predominant alpha with his eyes open, while fully awake, and fully conscious, and fully mobile. The disturbance or replacement of the predominant alpha frequency is called "alpha block" and is seen as a low voltage, higher frequency pattern. This is characteristic of an attentive state or alerting response. Although alpha rhythm responds sensitively to a number of features of CNS function, so many factors are involved that no simple unambiguous conclusion can usually be drawn from the presence or absence of alpha rhythm alone.

Generally, increased cortical arousal is associated with an EEG of lower voltage and higher frequency. Sleep or certain drugs act to produce a slowing of EEG frequencies, as does training and the controlled relaxation response of Jacobsen. In deep sleep, three waves per second are seen, the so-called delta waves. In moderate sleep, sleep spindles or bursts of 14 per second waves occur. Dreaming is associated with rapid eye movements (REM) and takes place in the range of drowsy to light sleep, the so-called emergent/Stage II type sleep. The term emergent is used to indicate a stage of sleep occurring following of period of one of the deeper stages of sleep. This phenomena will take place periodically through the night with many individuals exhibiting a particular sleep pattern unique to them alone. No dream activity is known to take place in delta sleep.

Additional studies are needed to establish a clear relationship between physiological measures, performance, and the level of CNS arousal in the drowsiness-extreme alertness continuum of wakefulness. In one study, it was observed that there were high levels of major period counts over ten second epochs during increased levels of arousal and lower major period counts during decreased periods of consciousness. The converse was generally true of the minor period count. Spectral analysis in this particular study showed quantitatively the shift of the major period modal band to slower frequencies and a shift of the minor period modal band to faster frequencies with decreasing arousal as sleep became deeper.

The combination of both frequency and amplitude domains in the activation analysis previously described, shows its sensitivity to some situations while indicating some ambiguity as a simple arousal indicator. Johnson and Ulett (50) examined 50 college students on three occasions using a modified EEG spectrum analyzer. Each subject was examined three successive occasions under quiet, eyes-closed conditions. Average values of the frequency spectrum for all students grouped by visits produced three curves of comparable frequency distribution; however, the curve corresponding to the first visit was approximately half the amplitude observed on subsequent visits. The authors concluded that the increased anxiety level of the subjects generated by their apprehension of the intial EEG examination produced this depression in amplitude at all frequencies. This shows that in this group of subjects, a decrease in anxiety was observable as a decrease in EEG amplitude at all observed frequencies between three and 33 cycles per second. The activation analysis which is sensitive to such amplitude changes may be a useful indicator of the anxiety level of an individual. The precise manner in which anxiety effects CNS function and the resultant level of performance is not yet clear, but is obviously a significant contributing factor in some stressful situations.

In a fatigue study at the USAF School of Aerospace Medicine, four pilots were required to complete a 24 hour simulator flight with only a two hour refueling stop in the middle of the run. An activation analysis of a continuous recorded EEG obtained on one of the flights showed a sustained high level of high frequency, low amplitude activity during the first several hours. This corresponded to the period of expressed anxiety on the part of the pilot as to his ability to perform adequately on the simulator. Interestingly, his first landing rated as one of the poorest of the eleven made during the 24 hours. Toward the end of the flight a generally lower level of activation level was observed with a marked tendency to fluctuate erratically between moderate and low levels. Thus, it is seen that this method of EEG analysis promises to contribute useful information regarding the general level and fluctuations of CNS arousal.

As we have indicated, it is probably a fair statement to make that at the moment there is little promise of new and exciting use of ongoing EEG material for the enhancement of pilot performance. In general, we can tell when a subject is getting drowsy, has gone to sleep, or, to a lesser degree of certainty, is simply inattentive. So we are left with inferring general state changes and its usefulness for monitoring the state of the organism. However, the event related potential called ERP (or Cortical Evoked Response) is another matter. Before discussing the ERP, the work of Donchin, et. al. has identified several interesting electronic signatures indicative of cortical activity (51). The first of these is called N200 and this electronic component is elicited whenever a rare or unexpected event occurs. Another of these is P300 and this endogenous component is seen in association with task revelant, rare stimuli. Another component is the contingent negative variation (CNV). This wave form is a slow negative shift of potential that occurs during the warned fore-period preceeding a motor or mental task. It begins very shortly after the warning stimulus and terminates after a response decision by the subject or the occurence of a stimulus which demands a response. The final, easily identified wave form is a readiness potential, the RP. This is similar to the CNV in that it is an event-proceeding negative shift. It is distinct from the CNV in a sense that it appears prior to self-paced voluntary responses. It's occurrence is independent of the presence of an eliciting or command stimulus. These endogenous components of the brain wave have been studied in connection with arousal, attention, selective attention, emotional valence, assessment of novelty, time estimation, uncertainty, detection of targets, differential identification of stimuli independent of size and shpae, and the semantic classification of linguistic symbols (52).

Electrocardiogram and Respiratory Indicators: In similar studies of performance the average values of ECG and respiration have consistently correlated well with general arousal level. The highest values are seen when performance demands and/or external stressors, such as threat or emergencies are introduced.

This relationship between average heart rate the level anxiety was nicely demonstrated in the highly significant series of experiments by Walter (53). Over a period of four years, he performed a series of complete defensive-avoidance conditioning procedures with 58 subjects, 37 normal and 21 psychiatric patients. He collected sufficient evidence in these experiments to distinguish two types of relations between average heart rate and blood pressure as indicated by pulse-wave velocity. One relationship showed a covariation in heart rate and blood pressure in the initial stages of excitement in normal subjects and in the cumulative effects of experimental stress in disturbed patients. This response was linked with other signs of generalized tension and anxiety and was associated with adaptive failure or confusion. The other type of response showed an inverse relationship, that is when heart rate increased, blood pressure fell. This was a transient effect showing blood pressure changes of about one-half the magnitude of the first type of response. This second type of response was frequently elicited by the penalty tone which was indictative of an erroneous response. This is an excellent example of increased resolution and reliability of interpretation afforded by the observation of simultaneous changes in two related physiological variables.

Ax (54) has reported a steady decrease in the mean value of the ratio of respiratory to nonrespiratory coupling in five subjects undergoing 123 hours of sleeplessness. This serves to indicate that the ratio of the length of time that the record is characterized by undisturbed respiration-coupled heart rhythm compared to cardioaccelerator reflex rhythm relates to CNS function under the stress of sleep deprevation. This is also an example of change in the peripheral expression of a central nervous reflex activity related to changes in levels of arousal.

Electrodermal Response Indicators: We have discussed two measures of electrodermal response, the BSR and the GSR, which are observed to change with the general level of CNS arousal. Levy (55) found that BSR compressed on a five centimeter per hour write-out was particularily valuable in monitoring states of consciousness. Under standard conditions he found the pattern of one individual's skin response to be consistently similar. However, the patterns of different subjects varied from an almost straight line to a wildly fluctuating one. The flat stable line which he obtained was consistently of low resistance due to frequent small amplitude GSRs. The more variable tracings were of higher average resistance showing less frequent, often large GSRs which tended to occur in groups. He also observed that persons who exhibited the low flat type of basic waking pattern seem to be able to maintain a more continuous and higher level of involvement in their environment than those persons showing a more variable tracing. In general, then, he reports a relatively stable, low value of BSR during aroused wakefulness, a more variable saw-tooth pattern during drowsiness, and a high resistance pattern during sleep.

Similar changes in BSR have been noted while monitoring pilots during flight. Here, resistance is initially low when the pilot starts flying the aircraft and gradually increases as he relaxes. His resistance drops if the co-pilot takes control of the plane and is lowest when the co-pilot is active in stall-type approach for landing. I suspect we would see the same response in a husband as his wife takes over driving down the turnpike.

In his series of conditioning procedures, Walter observed that an abundance of nonspecific GSRs was associated with muscular tension, slight tachycardia, raised blood pressure, and some EEG irregularities making up the familiar syndrome of tension/anxiety which constitutes one form of CNS arousal. These and other studies all report similar findings which indicate that the BSR, the number of specific GSRs, and the amplitude of specific responses when properly interpreted can indicate the general level of CNS arousal.

Individual Central Nervous System Responses: Having considered indicators of general CNS activity, we need to turn to individual CNS responses, since a significant portion of CNS activity concerns reflex responses to stimuli. Many reflex responses are sufficiently complex to involve a major portion of the suprasegmental CNS. The qualitative and quantitive identification of ongoing reflex response patterns should contribute greatly to an understanding of the functional status of the CNS at a particular time. Those CNS reflexes which have been identified include the adaptive reflex connected with the direction of a change of stimulus, the defensive reflex in response to a stimulus too strong for normal functioning, and the reflex responses *per se*, much evidence concerning central nervous system function can be gardnered from patterns of evoked responses and contingent effects. These latter responses require a specific applied stimulus of which the subject is aware and which tends to be distracting or alerting. Evoked responses may provide valuable guidelines for the interpretation of reflex response patterns observed in stressful situations or response patterns disturbed by specific activity in the environment.

The Orienting Reflex: The orientating reflex is of particular interest. This reflex, first identified by Pavlov, has been the focus of an extensive research program in the Soviet Union and has been the subject of many annual conferences. This reflex is characterized as an unspecific response initiated by any increase, decrease, or qualitative change of a stimulus independent of its modality. It is really the "what is it?" reflex of the central nervous system. It only acts to alert and prepare the individual for action. It does not itself initiate any action and is subject to extinction or habituation quite easily by repeated presentation of the same stimuli.

Two forms of the orienting reflex have been identified: (1) a generalized orienting and, (2) a localized orienting. For example, the initial presentation of a tactile stimulus produces a generalized response including an alpha block in the occipital and motor regions of the cortex, a GSR, an increase in muscle tension via EMG measurement, an eye movement, and a respiratory pause. After a few dozen representations, the only response which may be observed would be a transient alpha block in the motor region of the ocrtex. Here, the other components of the reaction have been inhibited, transforming the original reflex picture to a localized or more specific reflex. The total general orienting reflex picture also includes increase in heart rate, vasoconstriction of finger vessels, and vasodilation of the hand vessels. It is interesting to note that when we change the total sensory input, by adding an additional stimulus to the now habituated specific response, the generalized response is once again elicited. This demonstrates the preadaptive, rather than the adaptive nature of the orienting reflex.

One component of judgement and alertness includes the degree to which the individual is asking questions of, and interacting with his environment. While the frequency and magnitude of orienting reflexes may provide valuable indications of this interaction, it is fair to state that it is presently difficult to differentiate these effects from the general level of CNS arousal. Theoretical considerations and some preliminary reports suggest that it may be possible to demonstrate greater specificity of individual CNS reflexes. At least this is the desired direction for further research which is aimed at permitting clear distinction of CNS arousal to fear, anger, curiosity, and so forth.

EEG Indicators: As we have noted, the identification of the cortical components of central nervous system reflexes have depended primarily upon observation of alpha block indicating cortical arousal. We have also noted the presence of distinctive frequency shifts with cortical arousal even when the initial cortical rhythm is other than alpha. However, unaided visual interruption, or other gross measures of the EEG, do not permit easy identification of these changes, and thus, increasing attention has been focused upon the various methods of examining the EEG in a more microscopic fashion.

The use of toposcopical analysis of EEG records shows interesting contingent effects of so-called "social" versus "defensive" conditioning of alpha rhythms. Walter was able to demonstrate divergent changes in alpha when a subject was performing a task in cooperation with the experimenter's instructions, that is social conditioning, as compared to being thrown on his own resources to solve problems posed by the experimenter, that is defensive conditioning.

We have explored the cardiac, respiratory and electrodermal indicators of individual CNS arousal and other activities, and their relationships to each other, and can now turn to a more specific type of electrocortical activity which promises to give us a great deal of information in assessing human mental processing activity.

While it is well known and accepted that the task of pilotage and airborne systems controllers has changed dramatically from "seat-of-the-pants" type flying to sophisticated monitoring, pattern recognition and decision making, we are yet unable to identify, much less quantify, such mental processes. Nevertheless, as we have indicated, recent research shows that certain mental acts are related to specific electronically identifiable wave forms as well as to changes in related physiological parameters. Since such factors as fatigue, workload and stress (physiologic as well as psychologic) affect mental performance, it is highly desirable to be able to identify and quantify such measures. The main thrust of this research is the identification of specific cortical responses or response patterns evoked by specific stimuli. These event-related potentials can be characterized as an EEG response wave form, having both positive and negative values, with certain amplitudes and specific latencies and duration times. In studying these potentials, a series of positive and negative deflections is averaged for a group of trials. This characteristic wave-form signature, elicited by a specific stimulus, can be conceptually and empirically divided into two categories. The earlier components, those occuring in the first 100 milliseconds or so, subsequent to the stimulus, are referred to as exogenous. These exogenous components, reflect characteristics intrinsic to the stimulus event itself, such as loudness, brightness, intensity or other psychophysical attributes. This activity is considered to represent the processing of sensory information. The latter components, up to perhaps 600 milliseconds beyond the stimulus, are considered to be endogenous. These endogenous components reflect cognitive processes and attributes of the stimulus deriving not from its physical properties, but from its task-revelent context. As Lawrence (56) in an unpublished paper states, "it is these latter components, reflecting aspects of performance potentially applicable to cockpit or crew station situations which are of primary interest."

Lawrence speculates on a computer controlled workload allocation in a pilotage situation wherein "the use of brainwaves for the automatic enhancement of warning effectiveness could occur in two ways. A computer could sense some deficit in an operator's state of being, or potential deficit (anticipating a crisis) by making inferences about operator state from a set of physiological information channels. The computer could also observe a lack of attention to a warning display, or other performance deficit, and take action to somehow stimulate the human operator. Here, it would make inferences about observed deficits in operator performance, probably from assessment of ERPs, or their lack, in response to warning signal displays used functionally as probes. There exists the potential for use of brain wave indicators of dangerous operator state or behavior in that the presence of theta can predict drowsiness and deterioration of performace and of course the sleeping state can be readily detected. The detection of undesirable levels of arousal, that is, inappropriately high or low, or undesirable emotional states can probably be enhanced through the use of EEG information in conjunction with that obtained through other physiological or behavioral channels. Attention to a display could be assessed by a steady state ERP technique. ERPs could also be used to distinguish nonresponse to a warning signal resulting from a purposeful decision to ignore it, from accidental nonrecognition. This way, the hypothesized computer-based system could refrain from repetition or intensification of information which the operator has already processed and to which he presumably will eventually respond. Using brain wave information, a computer could also determine the occurance of an event like target acquisition and utilize this information better by employing a built-in control loop than a human observer could by employing a gross skeletal response such as pushing a button." It must be recognized with the advances of high speed, high performance aircraft there are circumstances which exist or will soon exist where a few milliseconds could provide an important advantage. This is especially true if one considers the small single savings in time and effort that would be accumulated over a rapidly successive series of events in a continuing recycling context of swift decision making and swift response.

As Lawrence points out, a more proximal goal would be the development of machine ability to sense such general intagibles as operator uncertainty and the need for more information or a need to maintain certain decision options and an upgrading of information relative to a particular pilot's role in an overall mission. Here, instantaneous, qualitative feedback to the machine could be given in the same way that varying intensities of temperature guide a missile toward a heat source. The ability to sense these variables continuously and sensitively would provide the basis for the very fine control of machine by man, perhaps even along the line of the creation of an artifically intelligent servomechanism so closely responsive in real time ot the operator's cognitions and perceptions that it could serve virtually as a functional extension of his own nervous system. It would seem that as we computer assist the functional machine we must also arrange to computer assist the functioning human being as the operator of that man-machine system. With this development, the problems of workload, performance, and stress would undoubtedly be resolved and laid to rest for once and for all.

REFERENCES

1. Chiles, W. D. Objective methods for developing indicies of pilot workload, FAA Report (FAA-AM-77-15), July 1977.

2. [illegible], R. E. The effect of binaural beats on performance, J. of Auditory Research, 3, 1961, pp. [illegible]-185.

3. [illegible], R. E., White, D. D. and Hartman, B. O. Neptune: a multielement task system for evaluating human performance, USAF School of Aerospace Medicine Technical Report (SAM-TR 69-25), Brooks AFB, TX, October 1969.

4. Trumbo, D. Instrumentation in Motor Skills Research, Amer. Psychol., 1969, 24(3), pp. 289-292.

5. Bahrick, H. P., Fitts, P. M. and Briggs, G. E. Learning curves--facts or artifacts, Psychol. Bull., 1957, 54, pp. 256-268.

6. Kennedy, R. S. and Coulter, X. B. Research note: the interactions among stress, vigilance, and task complexity, Human Factors, 1975, 17(1), pp. 106-109.

7. Demaio, J., Parkinson, S., Leshowitz, B. and Crosby, T. Visual scanning: comparisons between student and instructor pilots, USAFHRL Technical Report, 1976, June, No. 76-10.

8. O'Donnell, R. D. Handbook of human performance measures. Unpublished working draft, USAF Institute of Technology, Wright-Patterson AFB, OH 1972.

9. Wood, G. A. Neuromuscular correlates of sensori-motor performance in normal and fatigued states, Dissert. Abs., order no. 75-16, 616, 230 pp.

10. Wilkinson, R. T. and Houghton, D, Portable four-choice reaction time test with magnetic tape memory, Behav. Research Methods and Instrument, 1975, 7(5), pp. 441-446.

11. Gaillard, A. W. and Sanders, A. F. Some effect of ACTH 4-10 on performance during a serial reaction task, Psychopharmacologia, 1975, 42(2), pp. 201-208.

12. Bartz, A. E. Peripheral detection and central task complexity, Human Factors, 1976, 18(1), pp. 63-70.

13. Salzman, L. F. and Jaques, NON, Heart rate and cardiac cycle effects on reaction time, Percep. and Motor Skills, 1976, 43(3, pt. 2), pp. 1315-1321.

14. Thackray, R. I., Bailey, J. P. and Touchstone, R. M. Physiclogical subjective, and performance correlates of reported boredom and monotony while performing a simulated radar control task, FAA Office of Aviation Medicine Reports, 1975, HO. 75-8.

15. Holt, W. R. and Brainard, E. C. Selective hyperthermia and reaction time, Percept. and Motor Skills, 1976, 43(2), pp. 375-382.

16. Singleton, W. T. Deterioration of performance on a short-term perceptual-motor task, Floyd and Welford, Symposium on Fatigue, 28:511, pp. 163-172.

17. Sharkey, B. J., McDonald, J. F. and Corbridge, L. G. Pulse rate and pulmonary ventilation as predictors of human energy cost, Ergonomics, 1966, 9(3), pp. 223-227.

18. Schwarz, J. J. and Ekkers, C. L. Task and load difficulties in directing and regulating a complex technical system, Mens en Orderneming, 1976, March-April Vol., 30(2), pp. 85-108.

19. Frankenhaeuser, M. and Johansson, G. Task demand is reflected in catecholomine excretion and heart rate, J. of Human Stress, 1976, 2(1), pp. 15-23.

20. Montgomery, G. K. Effect of performance evaluation and anxiety on cardiac response in anticipation of difficult problem solving, Psychophysiologia, 1977, 14(3), pp. 251-257.

21. Nicholson, A. N., Hill, L. E., Borland, R. G. and Ferres, N. M. Activity of the nervous system during the let-down, approach and landing: a study of short duration high workload, Aerosp. Med., April 1970.

22. Kahneman, D., Tursky, B., Shapiro, D. and Crider, A. Pupillary, heart-rate, and skin resistance changes during a mental task, J. of Exp. Psychol., 1969, 79(1), pp. 164-167.

23. Kahneman, D. and Beatty, J. Pupil diameter and load on memory, Science, 1966, 154, pp. 1583-1585.

24. Dukes-Dobos, F. N. Fatigue from the point of view of urinary metabolites, Methodology in Human Fatigue Assessment, Haskimoto, Kogi and Grandjean, eds., Taylor and Francis, London, 1971.

25. Von Euler, U. S. Adrenalin and nonadrenalin in various kinds of stress. Symposium on Stress, Washington, D. C. Army Medical Service Graduate School and Walter Reed Army Service Center, 1953.

26. Hale, H. B., Williams, E. W. and Buckley, C. J. Aerospace aspects of the first non-stop transatlantic helicopter flight, Aerosp. Med., 1969, 40, pp. 718-723.

27. Selye, H. Stress, Acta Inc. Medical Publishers, Montreal, 1950.

28. Shannon, I. L., Pregmoir, J. R. and Brooks, R. A. Glucose concentrations in parotid fluid and blood serum following intravenous glucose loading, Oral Surg., 13:1010, 1960.

29. Warren, B. H., Ware, R. W., Shannon, I. L. and Leverett, S. D. Determination of inflight biochemical responses utilizing the parotid fluid collection technic, Aerosp. Med., August 1966, p. 796.

30. Basmajian, J. V. Muscles alive: their function revealed by electromyography, The Williams and Williams Co., Baltimore, 1974.

31. Lafevers, E. W. Power spectral density analysis of the electromyogram from a work task performed in a full pressure suit, Dissert. Abs., 1974, no. 75-1033, 82 pp.

32. Grandjean, E. and Kozi, K. Introductory remarks, Methodology in Human Fatigue Assessment, Haskimoto, Kogi and Grandjean, eds., Taylor and Francis, London, 1971.

33. Wolf, G. Construct validation of measures of three kinds of experimental fatigue, Percept. and Motor Skills, 1967, 24, pp. 1067-1076.

34. Saito, Y., Kogi, K. and Kashiwagi, S. Fractors underlying subjective feelings of fatigue, J. of the Science of Labor., 1970, 46, pp. 205-224.

35. Hess, W. R. Die funktionelle organization des vegetativen hervensystems, Basle: Benno Schwabe, 1948,

36. Kogi, K. and Saito, Y. A factor analytic study of phase discrimination in mental fatigue, Methodology in Human Fatigue Assessment, Haskimoto Kogi and Grandjean, eds., Taylor and Francis, London, 1971.

37. Ettema, J. H. and Zielhuis, R. L. Physiological parameters of mental load, Methodology in Human Fatigue Assessment, Haskimoto, Kogi and Grandjean, eds., Taylor and Francis, London, 1971.

38. Kashiwagi, S. Psychological rating of human fatigue, Methodology in Human Fatigue Assessment, Haskimoto, Kogi and Grandjean, eds., Taylor and Francis, London, 1971.

39. Storm, W. F. Hapenney, J. D. Mission-crew fatigue during rivet joint operations, USAF School of Aerospace Medicine TR-76-36, 1976.

40. McKenzie, R. E. A systems task used in the stress testing of special mission personnel, Human Factors, December 1965.

41. Welch, R. B., Longley, E. O. and Lomaev, O. The measurement of fatigue in hot working conditions, Methology in Human Fatigue Assessment, Haskimoto, [illegible] and Grandjean, eds., Taylor and Francis, London, 1971.

42. Grivel F. The influence of ambient and body heat on human without important physical load: II. specific heat stress effects evidenced in laboratory studies since 1958, Travail Humain, 1975, 38(2), pp. 223-244.

43. Bartlett, F. Psychological criteria of fatigue, Symposium on Fatigue, W. F. Floyd and A. T. Welford, eds., H. K. Lewis & Co., London, 1953, pp. 1-5.

44. Burch, N. R. and Greiner, T. H. A bioelectric scale of human alertness: concurrent recordings of the EEG and GSR, Psychiat. Res. Rep. Amer. Psychiat. Assn., 12:183-93, January 1960.

45. Luby, E. D., Grisell, J. L., Frohman, C. E., Lees, H., Cohen, B. D. and Gottliebb, J. S. Biochemical, psychological, and behavioral responses to sleep deprivation, Am. N.Y. Acad. of Sci., 96:71-9, 13 January 1962.

46. Burch, H. R. Automatic analysis of the electroencephalogram: a review and classification of systems, Electroenceph. Clin. Neurophysiol., 11:827-34, November 1959.

47. Riehl, J. L. Analog analysis of EEG activity, Aerosp. Med, 32:1101-8, December 1961.

48. Heymans, C. Reflexogenic areas of cardiovascular system, Perspect. Biol. Med., 3:409-17, Spring 1960.

49. McKenzie, R. E., Buckley, E. P. and Solaranis, K. An exploratory study of psychophysiological measurements as indicators of air traffic control sectors workload, Federal Aviation Agency Memorandum Report #157-524-03R., Washington, D. C., June 1966.

50. Johnson, L. C., Ulett, G. A., Sines, J. O. and Stern, J. A. Cortical activity and cognitive functioning, USAF School of Aerospace Med., 60:75:1-14, October 1960.

51. Donchin, E., McCarthy, G. and Kutas, M. Electroencephalographic investigations of hemispheric specialization, Language and hemispheric specialization in man: cerebral ERP's, Prog. Clin. Neurophysir., 3, J. E. Desmedt, ed., Karger:Basel:1977.

52. John, E. R. and Schwartz, E. L. The neurophysiology of information processing and cognition, Ann. Rev. Psychol, 1978, 29, Pato Alto: Annual Reviews Inc.

53. Walter, D. O., Advances in EEG analysis, Electroencephol. and Clin, Neurophysiology, Supplement No. 27, 1966.

54. Ax, A. and Luby, E. D. Autonomic responses to sleep deprivation, AMA Arch. Gen. Psychiat., 4:55-9, January 1961.

55. Levy, E. Z., Johnson, G. E., Serrano, J. Jr., Thaler, V. H. and Ruff, G. E. The use of skin resistence to monitor states of consciousness, Aerosp. Med. 32:60-6, January 1961.

56. Lawrence, G. H. Brain waves and the enhancement of pilot performance. Unpublished manuscript prepared for the Environmental Physiology Program, Office of Naval Research, Washington, D. C., June 1978 (Submitted to AGARD AMP WG-08).

ACKNOWLEDGEMENTS

I wish to express my appreciation for the cooperation of all of our authors, especially Doctors Gartner, Murphy, Buckley and Captain Perelli who gave me permission to abstract from their basic works. I hope that any shortcomings or criticism of these particular chapters will be directed to the editor and not the original sources.

I also wish to thank Diana L. Deyalc (A1C) and Robin G. Chavez (A1C) who did much of the library research required for Chapter 12.

Finally a special thank you for the secretarial services of Joyce Keller, Jeanette Jonietz, and Debra Coronado who all shared in a difficult task.

I hope our readers and the members of AGARD AMP WG-08 will recognize the outstanding support in personnel and services given by the United States Air Force School of Aerospace Medicine.

SUMMARY

The first three chapters provide a conceptual framework for workload, fatigue, and stress, within which to evaluate the remainder of this report. In each case, the authors attempted to be brief, to present a "capsule" statement of different definitions and orientations, and to the extent possible to prevent their own biases from entering into the text. What is the probability that all readers will be fully satisfied with the contents of the three chapters? Probably minimal, but hopefully few readers will be grossly dissatisfied.

The next three chapters, taken as a single unit, give a picture of the workload arena in a broad sense, partly historical and partly in terms of specific sub-problems and suggestions regarding selected methods or measurements. These chapters, therefore, augment the conceptual framework provided by the first three chapters.

Chapters 7, 8, and 9 come to grips in a concrete way with the critical issue of the anatomy of workload measurement technology. Chapter 7 provides a schema (a generalized representation on framework of a topic or problem area derived through an analytic but pragmatic process) for workload research. Chapter 8 describes a moderately less encompassing but still global program design applied to workload problems by one laboratory. Chapter 9 presents one modelling approach to workload--there are others, of course, as the author points out. As an aside, Chapter 9 is also a "preview" of an AGARDograph which the Aerospace Medical Panel is considering sponsoring in the near future. These three chapters are recommended particularly to laboratory directors, program directors, and supervisory scientists as tools for evaluation and goal-setting in their own programs in workload research.

Chapters 10 through 18 deal with selected measures applied to specific problems in specified settings. The first six are concerned with aircrew studies and the last three with air traffic control studies. There are many such sets of studies which could have appeared in this part of this report. These appear because they were offered and because we, the editors, valued both the investigator and the work he reported. In each case, the reader will be able to see how one investigation approached one specific problem using his own skills and the resources available to him. The virtue of this is that it lets the reader move from "frameworks," "schemas," etc., to concrete examples.

Chapter 19 stands by itself in this document. It is a modest compendium in which some measures from some domains (e.g., psychophysiology) are described and critiqued, the critiques clearly influenced by the skills, experiences, and biases of the author. The term "modest" is used to make a point. A compendium like this could be a probably useful and certainly very long handbook--probably two or three weighty volumes. The working group initially tasked itself with this objective, proposing to use a draft handbook offered by a US colleague, but it became apparent very early that the task was beyond the working group's capabilities (time, level of effort, etc.). This might be a useful future task for Aerospace Medical Panel sponsorship, though probably not in the conventional working group mode of operation.

Two points should be made in concluding. First, all papers after Chapter 4 contain pieces of studies, some data, analyses, findings, and so forth. The editors believe this enriches the more global parts of each chapter. Second, there are references given at the end of each chapter. Taken together--as a package--these constitute a highly useful bibliography.

Workload measurement methodology is in a continuing process of unfolding, acquiring new technology and instrumentation, moving into new measurement domains. This topic should be revisited by AGARD after a reasonable period of rest and recovery.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-AG-246	ISBN 92-835-1332-0	UNCLASSIFIED

5. Originator Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly sur Seine, France

6. Title

SURVEY OF METHODS TO ASSESS WORKLOAD

7. Presented at

8. Author(s)/Editor(s)	9. Date
Edited by B.O. Hartman* and R.E. McKenzie†	August 1979

10. Author's/Editor's Address	11. Pages
*Crew Technology Division, USAF School of Aerospace Medicine, Brooks Air Force Base, Texas, 78235, USA †San Antonio, Texas, USA	172

12. Distribution Statement This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.

13. Keywords/Descriptors

Military aircraft
Pilots (personnel)
Work measurement
Performance evaluation
Work sampling
Motor reactions
Senses

14. Abstract

Military aircraft are becoming increasingly complex, the associated avionics systems more sophisticated, and the mission profiles more demanding. The problem is to establish if such an increase in aircrew workload has become a limiting factor in the operational employment of some aircraft and to select valuable methods to assess it.

The measurement domain has been broken down into sensory threshold function tests, motor function and responses to psycho, physio and chemical excitation. The methodology includes a wide range of instrumentation, laboratories, inflight measurement and modelling methods, with the goal of compiling systematically and evaluating the multiplicity of approaches and techniques implied.

This preliminary survey is followed by a companion document (AGARD Advisory Report 139 – AR-139) where conclusions are set forth as far as workload measurement methodology is concerned.

This AGARDograph was sponsored by the AGARD Aerospace Medical Panel.

AGARDograph No. 246
Advisory Group for Aerospace Research and Development, NATO
SURVEY OF METHODS TO ASSESS WORKLOAD
Edited by B.O. Hartman and R.E. McKenzie
Published August 1979
172 pages

Military aircraft are becoming increasingly complex, the associated avionics systems more sophisticated, and the mission profiles more demanding. The problem is to establish if such an increase in aircrew workload has become a limiting factor in the operational employment of some aircraft and to select valuable methods to assess it.

P.T.O.

AGARD-AG-246

Military aircraft
Pilots (personnel)
Work measurement
Performance evaluation
Work sampling
Motor reactions
Senses

AGARDograph No. 246
Advisory Group for Aerospace Research and Development, NATO
SURVEY OF METHODS TO ASSESS WORKLOAD
Edited by B.O. Hartman and R.E. McKenzie
Published August 1979
172 pages

Military aircraft are becoming increasingly complex, the associated avionics systems more sophisticated, and the mission profiles more demanding. The problem is to establish if such an increase in aircrew workload has become a limiting factor in the operational employment of some aircraft and to select valuable methods to assess it.

P.T.O.

AGARD-AG-246

Military aircraft
Pilots (personnel)
Work measurement
Performance evaluation
Work sampling
Motor reactions
Senses

AGARDograph No. 246
Advisory Group for Aerospace Research and Development, NATO
SURVEY OF METHODS TO ASSESS WORKLOAD
Edited by B.O. Hartman and R.E. McKenzie
Published August 1979
172 pages

Military aircraft are becoming increasingly complex, the associated avionics systems more sophisticated, and the mission profiles more demanding. The problem is to establish if such an increase in aircrew workload has become a limiting factor in the operational employment of some aircraft and to select valuable methods to assess it.

P.T.O.

AGARD-AG-246

Military aircraft
Pilots (personnel)
Work measurement
Performance evaluation
Work sampling
Motor reactions
Senses

AGARDograph No. 246
Advisory Group for Aerospace Research and Development, NATO
SURVEY OF METHODS TO ASSESS WORKLOAD
Edited by B.O. Hartman and R.E. McKenzie
Published August 1979
172 pages

Military aircraft are becoming increasingly complex, the associated avionics systems more sophisticated, and the mission profiles more demanding. The problem is to establish if such an increase in aircrew workload has become a limiting factor in the operational employment of some aircraft and to select valuable methods to assess it.

P.T.O.

AGARD-AG-246

Military aircraft
Pilots (personnel)
Work measurement
Performance evaluation
Work sampling
Motor reactions
Senses

The measurement domain has been broken down into sensory threshold function tests, motor function and responses to psycho, physio and chemical excitation. The methodology includes a wide range of instrumentation, laboratories, inflight measurement and modelling methods, with the goal of compiling systematically and evaluating the multiplicity of approaches and techniques implied.

This preliminary survey is followed by a companion document (AGARD Advisory Report 139 – AR-139) where conclusions are set forth as far as workload measurement methodology is concerned.

This AGARDograph was sponsored by the AGARD Aerospace Medical Panel.

ISBN 92-835-1332-0

The measurement domain has been broken down into sensory threshold function tests, motor function and responses to psycho, physio and chemical excitation. The methodology includes a wide range of instrumentation, laboratories, inflight measurement and modelling methods, with the goal of compiling systematically and evaluating the multiplicity of approaches and techniques implied.

This preliminary survey is followed by a companion document (AGARD Advisory Report 139 – AR-139) where conclusions are set forth as far as workload measurement methodology is concerned.

This AGARDograph was sponsored by the AGARD Aerospace Medical Panel.

ISBN 92-835-1332-0

The measurement domain has been broken down into sensory threshold function tests, motor function and responses to psycho, physio and chemical excitation. The methodology includes a wide range of instrumentation, laboratories, inflight measurement and modelling methods, with the goal of compiling systematically and evaluating the multiplicity of approaches and techniques implied.

This preliminary survey is followed by a companion document (AGARD Advisory Report 139 – AR-139) where conclusions are set forth as far as workload measurement methodology is concerned.

This AGARDograph was sponsored by the AGARD Aerospace Medical Panel.

ISBN 92-835-1332-0

The measurement domain has been broken down into sensory threshold function tests, motor function and responses to psycho, physio and chemical excitation. The methodology includes a wide range of instrumentation, laboratories, inflight measurement and modelling methods, with the goal of compiling systematically and evaluating the multiplicity of approaches and techniques implied.

This preliminary survey is followed by a companion document (AGARD Advisory Report 139 – AR-139) where conclusions are set forth as far as workload measurement methodology is concerned.

This AGARDograph was sponsored by the AGARD Aerospace Medical Panel.

ISBN 92-835-1332-0

AGARDograph No. 246
Advisory Group for Aerospace Research and Development, NATO
SURVEY OF METHODS TO ASSESS WORKLOAD
Edited by B.O. Hartman and R.E. McKenzie
Published August 1979
172 pages

Military aircraft are becoming increasingly complex, the associated avionics systems more sophisticated, and the mission profiles more demanding. The problem is to establish if such an increase in aircrew workload has become a limiting factor in the operational employment of some aircraft and to select valuable methods to assess it.

P.T.O.

AGARD-AG-246

Military aircraft
Pilots (personnel)
Work measurement
Performance evaluation
Work sampling
Motor reactions
Senses

AGARDograph No. 246
Advisory Group for Aerospace Research and Development, NATO
SURVEY OF METHODS TO ASSESS WORKLOAD
Edited by B.O. Hartman and R.E. McKenzie
Published August 1979
172 pages

Military aircraft are becoming increasingly complex, the associated avionics systems more sophisticated, and the mission profiles more demanding. The problem is to establish if such an increase in aircrew workload has become a limiting factor in the operational employment of some aircraft and to select valuable methods to assess it.

P.T.O.

AGARD-AG-246

Military aircraft
Pilots (personnel)
Work measurement
Performance evaluation
Work sampling
Motor reactions
Senses

AGARDograph No. 246
Advisory Group for Aerospace Research and Development, NATO
SURVEY OF METHODS TO ASSESS WORKLOAD
Edited by B.O. Hartman and R.E. McKenzie
Published August 1979
172 pages

Military aircraft are becoming increasingly complex, the associated avionics systems more sophisticated, and the mission profiles more demanding. The problem is to establish if such an increase in aircrew workload has become a limiting factor in the operational employment of some aircraft and to select valuable methods to assess it.

P.T.O.

AGARD-AG-246

Military aircraft
Pilots (personnel)
Work measurement
Performance evaluation
Work sampling
Motor reactions
Senses

AGARDograph No. 246
Advisory Group for Aerospace Research and Development, NATO
SURVEY OF METHODS TO ASSESS WORKLOAD
Edited by B.O. Hartman and R.E. McKenzie
Published August 1979
172 pages

Military aircraft are becoming increasingly complex, the associated avionics systems more sophisticated, and the mission profiles more demanding. The problem is to establish if such an increase in aircrew workload has become a limiting factor in the operational employment of some aircraft and to select valuable methods to assess it.

P.T.O.

AGARD-AG-246

Military aircraft
Pilots (personnel)
Work measurement
Performance evaluation
Work sampling
Motor reactions
Senses

AGARD

NATO OTAN

7 RUE ANCELLE · 92200 NEUILLY-SUR-SEINE

FRANCE

Telephone 745.08.10 · Telex 610176

DISTRIBUTION OF UNCLASSIFIED AGARD PUBLICATIONS

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres. Further copies are sometimes available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Purchase Agencies listed below.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD – VSL
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Defence Scientific Information Service
Department of National Defence
Ottawa, Ontario K1A 0Z2

DENMARK
Danish Defence Research Board
Østerbrogades Kaserne
Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92 Châtillon sous Bagneux

GERMANY
Zentralstelle für Luft- und Raumfahrt-
dokumentation und -information
c/o Fachinformationszentrum Energie,
Physik, Mathematik GmbH
Kernforschungszentrum
7514 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force General Staff
Research and Development Directorate
Holargos, Athens, Greece

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3, Piazzale Adenauer
Roma/EUR

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 126
Delft

NORWAY
Norwegian Defence Research Establishment
Main Library
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Direcção do Serviço de Material
da Forca Aerea
Rua da Escola Politecnica 42
Lisboa
Attn: AGARD National Delegate

TURKEY
Department of Research and Development (ARGE)
Ministry of National Defence, Ankara

UNITED KINGDOM
Defence Research Information Centre
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RE

UNITED STATES
National Aeronautics and Space Administration (NASA)
Langley Field, Virginia 23365
Attn: Report Distribution and Storage Unit

THE UNITED STATES NATIONAL DISTRIBUTION CENTRE (NASA) DOES NOT HOLD STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

PURCHASE AGENCIES

Microfiche or Photocopy
National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22161, USA

Microfiche
Space Documentation Service
European Space Agency
10, rue Mario Nikis
75015 Paris, France

Microfiche
Technology Reports
Centre (DTI)
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RF
England

Requests for microfiche or photocopies of AGARD documents should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Facility
Post Office Box 8757
Baltimore/Washington International Airport
Maryland 21240; USA

Government Reports Announcements (GRA)
published by the National Technical
Information Services, Springfield
Virginia 22161, USA

Printed by Technical Editing and Reproduction Ltd
Harford House, 7-9 Charlotte St, London W1P 1HD