AD-A051 587 ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT--ETC F/G 5/9

ASSESSING PILOT WORKLOAD,(U)

FEB 78 W D CHILES, G A ELLIS, A H ROSCOE

UNCLASSIFIED AGARD-AG-233 NL

1 OF 1

AD A051587

END

DATE FILMED

4-78

DDC

AGARD-AG-233

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AGARDograph No. 233

Assessing Pilot Workload

NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY
ON BACK COVER

AGARD-AG-233

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARDograph No.233

ASSESSING PILOT WORKLOAD

This AGARDograph was prepared at the request of the Flight Mechanics Panel of AGARD.

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 Program and the Aerospace Applications Studies Program. 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.

Published February 1978

ISBN 92-835-74-X

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

PREFACE

The subject of cockpit workload is an important one for pilots and engineers, especially if they are concerned with evaluating handling qualities and guidance and display systems. This AGARDograph is mainly for such people.

It is not the intention of the authors to write a comprehensive and authoritative book on workload; it would be presumptuous to think we could do so. Even to deal adequately with all the aspects would be impossible and only short term workload will be considered here.

The AGARDograph contains relatively little philosophical discussion although various ideas and definitions of the term "pilot workload" are introduced in Chapter 1. In addition, the author of each chapter discusses briefly his own idea of what is meant by workload. It will be apparent that although there are many interpretations and definitions of pilot workload there are only two broad conceptual areas. The first considers workload in terms of the demands of the flight task, the second conceptual area is directed to workload as the effort required of the pilot to satisfy these demands. In general, estimation of workload based on task-related concepts results in theoretical values whereas estimation based on response-related concepts results in levels of actual workload.

This is a fundamental difference which is difficult, perhaps impossible, to resolve; it is the main reason why there is no single acceptable definition.

In-flight assessment of pilot workload depends largely on the measurement of pilot effort in one form or another, and the contents of this volume tend to be slanted in that direction. Subjective and physiological methods, reviewed in Chapters 2 and 3 respectively, are particularly relevant. Objective methods, discussed in Chapter 4, contain techniques appropriate to workload both as pilot effort and as task demands. Data from these latter techniques are expecially useful for constructing models and for predicting levels of workload.

The use of modelling techniques to estimate values of theoretical workload will be considered in a proposed supplement to this AGARDograph entitled Engineering Methods.

FOREWORD

In Spring 1973 the FMP discussed the possibility of writing and publishing an AGARDograph on "Assessing Pilot Workload". It was decided this should be undertaken in collaboration with the ASMP. A sub-committee, with members of both panels, was established and the terms of reference were drawn up.

The unusually long delay between the start of the work and the publication of the results highlights the difficulties the panel has had to overcome. There has been universal discussion of what constitutes "pilot workload" and many authors have written papers on their favourite concept of workload but there are apparently very few who are able and willing to collate these papers to make them useable and understandable by the pilots and flight test engineers to whom this AGARDograph is addressed.

As a consequence it proved difficult to find suitable authors, a problem made worse by three authors of particular sections having to withdraw at various stages because of changes in their primary commitments.

The result, as it is presented here, is an attempt to review the work done in the western world on the subject of "pilot workload" and to draw preliminary conclusions. Criticism may still be justified in that the work is incomplete and somewhat inconclusive. Every effort has been made to refer to relevant published work on the subject, but, in view of the very great number of papers some – perhaps important – work may have been overlooked. The interpretation of workload given here will not satisfy everyone. It should be borne in mind, however, that a subject whose title still defies a commonly agreed definition may, nevertheless, be well served by this preliminary interpretation.

As a result of its collaboration in this task the ASMP initiated a number of activities – in particular one described in AGARD CP 216 which, similarly, did not produce a generally applicable method for measuring workload.

Since there is obviously no immediate prospect of a "break-through" that would completely eliminate all confusion and controversy, it is the FMP's view that the work should be published without further delay, as it stands.

If the AGARDograph does no more than stimulate informed and constructive criticism of current ideas, a large part of the purpose of this effort will have been achieved.

On behalf of the FMP, I would like to thank the authors of the different sections and, in particular, the editor Dr. Roscoe, for their work. Special thanks must also be given to the FMP coordinator for this AGARDograph, Mr D.Lean, whose unrelenting efforts achieved the realization of the FMP's intentions and to the reviewers of the volume: Professor Doetsch, Mr J.Renaudie and Dr. I.C.Statler, for their positive criticism and helpful suggestions.

HEINZ MAX
Chairman, Flight Mechanics Panel

The FMP learned with regret that Dr. Dean Chiles of the USA, Author of Chapter 4, died shortly before this publication went to press.

CONTENTS

CHAPTER 1

INTRODUCTION

by

Alan H.Roscoe
Royal Aircraft Establishment
Bedford, England MK41 6AE

CONTENTS

INTRODUCTION

1.1 PILOT WORKLOAD

Flying an aeroplane imposes a load on the pilot who has to expend an amount of physical and mental effort to accomplish the task. This simple statement belies the difficulty of defining pilot workload and a review of the literature highlights the diversity of interpretation and the vagueness which exists.

There is no one acceptable definition but several authors have identified effort as the main theme in their concept of workload. In their handling qualities rating scale, Cooper and Harper[1] ask "Is adequate performance attainable with a tolerable pilot workload?" and they defined pilot workload as "the integrated physical and mental effort required to perform a specified piloting task". Tennstedt[2] described pilot workload as "a summation of such processes as perception, evaluation, decision making and actions taken to accommodate those needs generated by influences originating within or without the aircraft". Jenney and his colleagues[3] defined workload as ". . . . the level of effort required to perform a given activity or complex of tasks".

The idea of workload as effort is one with which many pilots would agree. It caters for the individual ways in which pilots respond to the demands of the flight task by allowing for such variables as natural ability, training, experience, age and fitness. However, there are other important aspects of the flight task which may be considered to be equally relevant in forming concepts of workload. They provide a fertile soil for controversy.

Following a conference on flight deck workload and pilot performance, Benson[4], in his technical evaluation, pointed out that ". . . . many of the papers presented emphasised the integrative nature of the workload concept". Jahns[5] likewise considered workload as an integrative concept but also found it practical to think of three functionally related attributes, namely: input load, operator effort, and work result. In a comprehensive survey of workload concepts, Gartner and Murphy[6] adopted Jahn's classification though with minor changes. They contemplated three notions based on: workload as a set of task demands, workload as effort, and workload as activity or accomplishment. Three variables were also considered by Billings and Lauber[7], these were: the demands of the task – requirements of the system; the effort put forth by the pilot – his workload; and the results of that effort – the performance of the system.

Jahns[5] in addition to considering workload as an integrative concept also suggested that in broad terms ". . . . workload is the extent to which an operator is occupied by a task". He went on to indicate that ". . . . this definition stems from the time-limited capability of the human operator". Brown et al[8] emphasised the time element in their definition: "Flight crew workload is the ratio of summation of required crew-equipment performance time to time available within the constraint regulated by a given flight or mission". The introduction of time ingredients into task demands is a major factor in formulating ideas on workload. This was highlighted by White[9] in a review of task analysis methods and mental workload, when he stated: "Time demands are important components of workload".

Cooper and Harper[1] refer, in their scale, to "pilot compensation" using the term to indicate that the pilot must increase his workload to improve aircraft performance. They also state that "it is the measure of additional pilot effort and attention required to maintain a given level of performance in the face of less favourable or deficient characteristics". The idea that a pilot has the ability to compensate implies that he has spare capacity; Clement and his colleagues[10] suggested a definition of pilot workload based on this notion, namely: ". . . . the ability (or capacity) to accomplish additional (expected or unexpected) tasks".

Although there are many different definitions and concepts of pilot workload it is generally acknowledged that there are two main areas for consideration, they are task-related and pilot-related aspects. In Gartner and Murphy's[6] classification of workload, task-related aspects are the task demands, and pilot-related aspects are effort and activity or accomplishment. These authors point out that demand-oriented expressions of workload are free of operator response or response capabilities; because of this they observe that "it would seem advisable to associate demand only with input or stimulus-oriented variables and to reserve workload for the response-oriented variables. Billings and Lauber[7] also differentiated ". . . . the demands placed on the man by his vehicle and the system from his response to those demands – his workload".

It may be useful to consider workload as a multifaceted concept, primary facets being formed by the three variables: demands of the flight task, pilot effort, and results. Minor or secondary facets can then be formed by the various methods used for assessing levels of workload. These will be largely dependent on the experience, discipline and interest of the investigator. It follows that any reference to pilot workload must identify the particular interpretation and the method used to assess levels.

1.2 CLASSIFICATION

It is customary to divide workload into physical and non-physical or mental components, though it is not always easy to identify a clear dividing line between them. Cooper and Harper[1] distinguished between physical and mental effort and Rolfe and Lindsay[11] stressed the importance of differentiating between physical and mental aspects of workload. Rolfe and his colleagues[12] divided pilot workload into three components: physical, perceptual, and mental, and described a simulator experiment in which they attempted to separate them.

1.2.1 Physical Workload

It is a relatively simple matter to assess pure physical activity by using accepted physiological measuring techniques to estimate the body's metabolism. The physical content of pilot workload, when compared with physical work in general, is usually quite low. Metabolic studies by Billings et al[13] and by Littell and Joy[14] have shown that the physical activity involved in flying helicopters and light fixed wing aircraft can be classed as sedentary or light work. Blix and his co-workers[15] assessed the metabolic effect on pilots flying helicopters and large transport aircraft and confirmed that the level of physical workload is low.

1.2.2 Mental Workload

A physical element is present in most flight tasks but it is the mental component, in particular, which causes so much confusion in forming concepts and definitions. In 1958 Cohen and Silverman[16] suggested that measurement of mental effort might include ". . . . evaluating the peripheral, integrative, and motoric abilities of the pilot, as well as emotional, physiologic and hormonal responses". Firth[17] pointed out that because of the complexity and covert nature of mental functions, such as information assimilation and decision making, there is a lack of knowledge about the nature of mental workload. Many studies of mental activities, especially information processing and decision making, have involved the construction of models which, by introducing feedback loops and neurophysiological components, have been made increasingly complex[5,18]. Experimental evidence supports the hypothesis that man has only a single channel capability for processing information and making decisions[19]. Based on this hypothesis is the idea of a maximum capacity for mental processing which, if exceeded by the demands of the task, would lead to overload and breakdown with a consequent deterioration in performance[20].

1.2.3 Duration of Workload

It is convenient to classify workload according to duration. Howitt[21] considered three timescales: 'immediate' workload which is that associated with a particular phase or sub-phase of flight, 'duty day' workload, and 'long term' workload which considers the effects of a sequence of working days over a specific duty period.

This AGARDograph is primarily concerned with immediate or short term workload, though it is important to bear in mind that long term workload is influenced by levels of workload generated by sub-phases of flight. Conversely, the effects of long term workload may modify pilot response to the immediate flight task.

Psychophysiological effects on aircrew, caused by long term workload associated with flights of different lengths, have been studied in detail. In 1958, Marchbanks[22] reported results of an investigation into the effects of a 22½ hour mission on the four man crew of a B52 bomber. Several studies on the different aspects of long duration missions have been carried out by Hale and his associates[23,24]. Howitt et al[25] observed crew activity, measured heart rate, and used biochemical techniques in a study of pilot workload in long-haul transport aircraft. An empirically derived model mission was used by Hartman and Cantrell[26] to investigate the effects of disruptions in sleeping, eating and working patterns. Mills and Nicholson[27] examined the relationship between workload and sleep patterns during a long range air-to-air refuelling exercise.

These and similar studies have shown the important and complex interrelationship between long term workload, fatigue, variable eating, sleeping and working habits, time zone changes and alterations in biological rhythms. Short term workload is more or less unaffected by such factors. It should be noted that though several measuring techniques are commonly used to assess both long and short term workload, methods having a long response time are not really suitable for estimating immediate workload.

1.3 INFLUENCE OF STRESS

Many authors have emphasised the mental stress component of pilot workload and its synergistic effect on the task. Stress produces physiological, psychological, and occasionally pathological effects on pilots known as strain[28,29]. A generally accepted definition of strain, based on the work of Selye[30], is "the non-specific response of the body to any demand made upon it". In the context of flying it is usual to divide stresses into two main types: those of environmental origin, often termed physical stresses, for example noise, vibration, abnormal temperatures and accelerations; and those of psychological origin such as fear, exhileration, and frustration. The effects of the former type are well recognised and documented[31] whereas psychological stresses are not easy to identify nor to describe. Responsibility and pacing are psychological stresses which are obviously part of workload but other stresses, in particular those of emotional origin, may be quite unrelated to the flight task.

It is difficult to estimate the effects of psychological stress on pilot performance and workload; most experimental work has been done in laboratories where it is almost impossible to create a realistic flight situation. Although risk and fear of physical harm have been cited by several authors as being a common flight stress, there is evidence to show that for the experienced pilot in current flying practice it is normally an insignificant factor[32,33].

1.4 WHY ASSESS PILOT WORKLOAD?

Rolfe and Lindsay[11] answered this question with one word: "reliability"; they explained their simple answer by pointing out that whereas the machine is becoming more reliable, man is looked on as the most suspect component of the system. They cited Senders[34] who stated that to a large extent the reliability of the man is a function of the load placed upon him.

The association between pilot workload and flight safety is indisputable and there have been many instances where abnormal flight deck workload levels have been implicated, directly or indirectly, as causative factors in aircraft accidents[35]. In most cases an overlead situation has been identified but there is evidence to suggest that low levels of workload have also been responsible for accidents. At present there is no confident way of predicting overload and subsequent breakdown of performance; thus behavioural scientists have a direct interest in assessing workload to identify limits and to derive estimates of reliability.

Although it is important to understand the manner in which human pilots respond to the demands of the flight task, this review is more concerned with levels of workload determined by the aircraft, systems and procedures, and with the effect of extraneous factors, such as weather, on these levels.

Before considering where improvements to handling qualities or systems may be beneficial it is necessary to get some idea of overall levels of workload for particular phases or sub-phases of flight. There is also a need to identify any peaks or troughs which may be present and which may be readily smoothed out.

Design engineers are generally aware of the problems associated with high levels of workload during the more demanding phases of flight, exemplified by the take-off and the approach and landing. This is typified by the design philosophies for two advanced medium STOL transports currently being developed. The Boeing YC-14 is to be fitted with an electronic flight control system" . . . . designed to minimize pilot workload during precision landings and to ensure that the aircraft handles as easily in the slow-speed, short take-off and landing mode as in cruise flight".[36] And the McDonnell Douglas YC-15 is to have an integrated flight control and augmentation system (IFACS) designed to reduce pilot workload during slow approaches to STOL landings[37].

Improvement in workload levels as a result of design changes may not always be reflected in improved performance; a pilot may adjust his effort according to the demands of the flight task without affecting performance. In 1956 Duddy[38] highlighted the difficulty of assessing pilot effort and cited an example where some measure of workload would have been of practical value: the advantage of a yaw damper in the weapon aiming of a directionally unstable fighter was not apparent as aiming accuracy was not improved by the damper. It was obvious that pilot workload was reduced, but by how much? Spyker and his colleagues[39] observed that: "An evaluation procedure which relies exclusively on performance measure is inadequate. That is, a pilot with one configuration may work twice as hard as he does with another, yet achieve equal performance for both".

Changes to aircraft handling qualities, displays and control systems designed to improve performance and to reduce workload may not always achieve the desired effect. For example, the use of autothrottle reduced pilot workload during curved landing approaches at Gibraltar in a HS Trident jet transport[40] but a poor system may well increase workload by causing frequent pitch and trim changes[41]. The addition of autostabilisation may improve handling but if the integrity of the system is low it may be necessary to increase the monitoring of the system itself, thereby leaving the workload level unchanged, or possibly increased[42]. Alterations in the display of information to the pilot can lead to changes in workload, but not always in the right direction; superfluous or ambiguous information may increase workload. Cooper[43] commented that ". . . . the need is not only to find a way to get more information into the cockpit, but to do it in a manner which neither compromises the existing pilot-aircraft performance nor increases the workload".

Having determined workload levels for normal flight conditions it is important to assess the effects of turbulence, poor visibility, and other extraneous factors.

As well as estimating individual levels it is sometimes necessary to assess the effects of varying the proportion of workload shared between different crew members. Nicholson et al[44] noted the advantages of shared workload during difficult landing approaches in a large transport aircraft. Provision of an extra crew member seems to be a logical way of reducing workload but as Wallick[45] pointed out ". . . . the addition of more crew members to reduce the individual workload is partially self-defeating since each additional member requires co-ordinations with existing members – thereby increasing the total flight deck workload". ter Braak[46] reached a similar conclusion after comparing workload levels for one and two crew members of a strike aircraft during simulated tactical missions.

1.5 ASSESSING PILOT WORKLOAD

Ideally, assessment or measurement of pilot workload should be objective and result in absolute values; at present this is not possible nor is there any evidence that this ideal will be realised in the foreseeable future. It is also unfortunate that the human pilot cannot be measured with the same degree of precision as can mechanical and electronic functions.

Methods used for assessing workload can be broadly divided into subjective, physiological, objective, and engineering techniques; their use almost invariably involves crossing interdisciplinary boundaries. The practical application of these techniques to the three conceptual areas considered earlier: the flight task, pilot effort, and performance, results in a measure of workload which will have a specific interpretation depending on the particular technique selected.

An important difference between assessments of pilot workload based on demand-oriented and effort-oriented concepts is that the former result, primarily, in levels of *theoretical* workload whereas the latter result in levels of *actual* workload. The measurement of performance *per se* may prove to be of little value in assessing workload as it may remain unchanged despite alterations in levels of workload caused by changes in demands or by variations in effort. Nevertheless, it is more or less essential to monitor performance when using techniques directed both to workload as demand and to workload as effort.

A study of the literature shows that the most widely used techniques for estimating levels of pilot workload are those based on effort related concepts. Subjective methods using some form of pilot opinion rating are particularly useful for in-flight assessment. These methods, which are related to those used for evaluating aircraft handling qualities, are discussed in Chapter 2. Subjective opinions are sometimes viewed with suspicion by engineers more familiar with measuring absolute values. However, though different techniques for subjectively rating workload may vary in their reliability, a well designed questionnaire combined with a rating scale is probably the best single measure of short term workload.

Physiological methods of estimating pilot workload are based on the concept of neurological arousal or activation. This is a state of activity in the nervous system which varies along a continuum of intensity from deep sleep at one end to hyperexcitability at the other. It has been shown that arousal and performance are related and that for a skilled and difficult task an optimum level of arousal is necessary to achieve maximum performance. Measures of arousal might, therefore, be expected to indicate levels of operator workload.

Physiological indices such as heart rate, muscle tension, respiration and so on, reflect the level of arousal. Theoretically there is a whole gamut of variables available to the life scientist interested in measuring a pilot's physiological response to the demands of the flight task; although only a few of these variables are suitable for routine use in aircraft. These are discussed in Chapter 3 which also reviews the wider range of techniques suitable for use in laboratory experiments.

Many studies, especially those carried out in laboratories and simulators, have tended to cast doubt on the value of physiological methods. Nonetheless, there is good evidence from a number of flight trials to support their use in assessing levels of workload for handling pilots during realistically demanding flight tasks.

Objective methods for assessing pilot workload can generally be divided into observational or analytical based techniques and measurements of performance with and without secondary or loading tasks. Analytical techniques, based on time-and-motion type studies, are particularly useful for assessing workload in laboratory cockpit mock-ups and in flight simulators, with a view to optimising designs and operational procedures. By using observational techniques it is possible to investigate various aspects of the flight task; for example, scanning or visual workload levels can be estimated by using eye point of regard monitors. Unfortunately, observational techniques do not reveal the true extent of the covert mental activity that forms such an important part of workload.

Measurement of operator performance as a means of estimating workload has been a technique used by many research workers in flight simulators and in aircraft. But, as observed earlier, performance frequently remains the same despite an obvious increase in task difficulty. According to Brown[47] "If man has reserve capacity the perceptual load imposed on him cannot be evaluated by measuring his performance on the system because he makes no errors – by adding a second task so that total information to be handled exceeds the man's total capacity, errors can be forcibly produced". Secondary or loading tasks are commonly applied to measuring mental load during complex tasks and in comparing specific designs or systems. But it is difficult and probably unrealistic to apply secondary task techniques, as used in the laboratory, to real-flight measurement.

Whereas subjective and physiological methods are appropriate only to effort-related concepts of workload, objective methods tend to be concerned with all three conceptual areas. Objective methods are described in Chapter 4 which also refers to some typical examples of their practical application.

The man-machine system is characterised by complex interactions and inter-relations which have considerable influence on pilot workload. A greater understanding of workload, and of pilot behaviour in general, has been acquired through the construction of various models of the man-machine interface. Early studies during the 1950s resulted in models based on the analysis of pilot control activity during simple tracking tasks. Since that time, subjective, objective, and physiological methods have been used to identify the demands of the flight task together with the pilot's response to those demands. McRuer[48] coined the phrase "dynamical dissection of the human" to describe the analytical techniques used to measure human responses. Development of techniques based on data derived from this kind of detailed analysis (see Chapter 4), and aided by the availability of advanced computer facilities, has led to the construction of highly sophisticated models[49,50]. Most of these have been based on human operator control dynamics, but the increasing tendency for pilots to become systems supervisors, rather than active controllers, has necessitated the introduction of models based on the pilot as a monitor and decision maker[51,52].

Several workers have applied modelling techniques, based on single and multi-loop situations, directly to the study of pilot workload[53,54]. Bernotat and Wanner[18] discussed workload in terms of a multi-loop model, and Clement and his colleagues [10] observed that the closed-loop theory for manual control display systems provides a rational basis for directing engineering analysis towards excess control capacity as ".... a predictable practical measure of pilot workload".

Modelling techniques are particularly attractive to engineers but perhaps a word of caution is necessary. For example, it tends to be assumed that the human pilot always behaves in an optimum and predictable manner whereas in practice, of course, this is not so. Christenson[55] made the point that: "This seems to be the age of models" and he continued, "A model is never the real thing; otherwise it wouldn't be called a model". However, Sheridan[56], in countering possible criticism of modelling human behaviour, argued: ".... that we are dealing with man-machine interactions which are quite utilitarian and mechanistic to begin with. Therefore, such mechanistic mathematical models have a face validity. Anyway, when the stimulus and response are well defined and the decision criteria straightforward, the models are useful because they are good predictors of the aspects of human behaviour which are important".

Mathematical modelling, using data derived from detailed analysis of well defined flight tasks, is an engineering technique with increasing potential for predicting levels of workload for new aircraft and systems.

Modelling techniques are discussed further in a proposed supplement to this volume entitled Engineering Measures; L.D.Reid considers mathematical models based on human operator describing functions and J-C Wanner presents a paper on the multi-loop concept of pilot workload.

To-date, most studies of workload have been done in laboratories and flight simulators and there is a noticeable lack of data obtained from the real-world. Jahns[5] noted this fact and suggested that as a result, techniques for assessing workload ".... tend to control or ignore the synergistic effect of tasks found in the operational environment". Laboratory or simulator experiments, and particularly the modelling techniques derived from them, 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 behaviour and hence his workload. Assessment of workload in simulators is important for developing methodologies and for initially evaluating new cockpits and systems. However, it is eventually necessary to obtain more information about levels of actual workload associated with different phases of *real flight*.

It should be noted that each group of methods discussed in this AGARDograph has an important place in the study and assessment of pilot workload. At present, though, it does not seem possible to combine the results of these different methods to produce an overall index of workload.

REFERENCES

1. Cooper, G.E., Harper, R.P. — *The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualitities* – NASA Tech Note TN D – 5153, Washington DC, 1969.

2. Tennstedt, C.R. — *The Pilot – His Problems and Requirements.* In: Symposium. Flight Deck Environment and Pilot Workload – Royal Aeronautical Society, London, 1973.

3. Jenney, L.L., Older, H.J., Cameron, B.J. — *Measurement of Operator Workload in an Information Processing Task.* NASA Contractor Report CR 2150, Washington DC, 1972.

4. Benson, A.J. — Conference Proceedings No.56 – *Measurement of Aircrew Performance.* AGARD, Paris, 1969.

5. Jahns, D.W. — *A Concept of Operator Workload in Manual Vehicle Operation.* Forschungsinstitut für Anthropotechnik, Forschungsbericht Nr 14, 1973.

6. Gartner, W.B., Murphy, M.R. — *Pilot Workload and Fatigue. A Critical Survery of Concepts and Assessment Techniques.* NASA Report No. TN D 8365, Washington DC, 1976.

7. Billings, C.E., Lauber, J.K. — *Short-term Workload in Airline Operations.* Paper presented to IATA Conference on 'Safety in Flight Operations'. Istanbul, 1975.

8. Brown, E.L., Stone, G., Pearce, W.E. — *Improved Cockpits Through Flight Crew Workload Measurement.* Paper presented to Second Advanced Aircrew Display Symposium. Naval Air Test Centre, Patuxent River, 1975.

9. White, R.T. — *Task Analysis Methods:* A Review of Development of Techniques for Analysing Mental Workload in Multiple-task Situations. McDonnell Douglas Corporation, Report No.MDC J. 5291, 1971.

10. Clement, W.F., McRuer, D.T., Klein, R.H. — *Systematic Manual Control Display Design.* In Conference Proceedings No.96. Guidance and Control Displays. AGARD, Paris, 1971.

11. Rolfe, J.M., Lindsay, S.J.E. — *Biological Measures of Workload.* In: Symposium. *Flight Deck Environment and Pilot Workload* – Royal Aeronautical Society, London, 1973.

12 Rolfe, J.M. et al. — *Evaluating Measures of Workload Using a Flight Simulator.* In Conference Proceedings CP 146. Simulation and Study of High Workload Operations. AGARD, Paris, 1974.

13. Billings, C.E., Bason, R., Gerke, R.J. — *Physiological Cost of Piloting Rotary Wing Aircraft.* Aerosp Med. 41, p.256, 1970.

14. Littell, D.E., Joy, R.J.T. — *Energy Cost of Piloting Fixed and Rotary-wing Aircraft.* J. App. Physiol. 26, pp.282-285, 1969.

15. Blix, A.S., Stromme, S.B., Ursin, H. — *Additional Heart Rate – An Indicator of Psychological Activation.* Aerosp Med. 45, pp.1219-1222, 1974.

16. Cohen, S.I., Silverman, A.J. — *Measurement of Pilot Mental Effort.* AGARD Report 148, Paris, 1957.

17. Firth Patricia, A. — *Psychological Factors Influencing the Relationship Between Cardiac Arrhythmia and Mental Load* – Ergonomics. 16, pp.5-16, 1973.

18. Bernotat, R.K., Wanner, J. — *Pilot Workload.* In Conference Proceedings No.106. Handling Qualities – AGARD, Paris, 1971.

19. Welford, A.T. — *Single-channel Operation in the Brain.* In: Attention and Performance. Sanders, A.F. (Ed) North Holland, Amsterdam, 1967.

20. Miller, J.G. — *Input Overload and Psychopathology.* Amer. J. Psychiat. 116, pp.695-704, 1960.

21. Howitt, J.S. — *Flight-deck Workload Studies in Civil Transport Aircraft.* In Conference Proceedings No.56. Measurement of Aircrew Performance. AGARD, Paris, 1969.

22. Marchbanks, V.H. — *Effect of Flying Stress on Urinary 17 – Hydroxycortiscosteriod Levels.* J. Aviat Med. 29, pp.676-682, 1958.

23. Hale, H.B. et al. — *Endocrine-metabolic Effect of Unusually Long or Frequent Missions C – 130E or C – 135B Aircraft.* Aerosp Med. 39, pp.561-570, 1968.

24. Hale, H.B. et al. — *Physiological Cost of Prolonged Double Crew Flights in C – 5 Aircraft.* Aerosp Med. 44, pp.999-1008, 1973.

25. Howitt, J.D. et al. — *A Preliminary Study of Flight Deck Activity in Civil Air Transport Aircraft.* Flying Personnel Research Committee Report No.1240, London, 1965.

26. Hartman, B.O., Cantrell, G.K. — *Sustained Pilot Performance Requires More than Skill.* Aerosp Med. 38, pp.801-803, 1967.

27. Mills, N.H., Nicholson, A.N. — *Long Range Air-to-air Refuelling. A Study of Duty and Sleep Patterns.* In Conference Proceedings No.146. Simulation and Study of High Workload Operations. AGARD, Paris, 1974.

28. Billings, C.E. et al. — *Stress and Strain in Student Helicopter Pilots.* Aerospace Med. 44, pp.1031-1035, 1973.

29. Hoffelt, W. Gerbert, K. *Some Critical Comments on the Measurement of In-flight Strains.* In Conference Proceedings No.101 – Performance and Biodynamics Stress. AGARD, Paris, 1972.

30. Selye, H. *Stress and Aerospace Medicine.* Aerospace Med. 44, pp.190-193, 1973.

31. Murray, R.H. McCally, M. *Combined Environmental Stresses.* In: Bioastronautics Data Book. Parker, J.F, West, V.R. (Eds) NASA, Washington DC, 1973.

32. Roman, J. *Risk and Responsibility as Factors Affecting Heart Rate in Pilots.* Flight Research Programme II. Aerospace Med. 36, pp.518-523, 1965.

33. Melton, C.E. et al. *Physiological Studies on Air Tanker Pilots Flying Fire Retardant Missions.* FAA Office of Aviation Medicine Report No. AM-68-26. Washington DC, 1968.

34. Senders, J.W. *The Estimation of Operator Workload in Complex Systems.* In: De Greene (Ed). Systems Psychology. McGraw-Hill, London, 1970.

35. Succhi, G.D. Sells, S.B. *Information Load and Three-man Flight Crews:* An Examination of the Traditional Organisation in Relation to Current and Developing Airlines. Aerospace Med. 40, pp.402-406, 1969.

36. Elson, B. *Eased Pilot Workload Sought in YC-14.* Aviation Week and Space Technology 102(8) pp.56-57, February 1975.

37. Fink, D.E. *Austerity Dictates Simple YC-15 Design.* Aviation Week and Space Technology. 103(9) pp.57-63, September 1975.

38. Duddy, R.R. *The Quantitative Evaluation of an Aircraft Control System.* AGARD Report No.29. NATO Paris, 1956.

39. Spyker, D.A. et al. *Development of Techniques for Measuring Pilot Workload.* NASA Contractor Report CR 1888. Washington DC, 1971.

40. Harper, P.J.A. *Pilot as an Aircraft Systems Manager.* Technical Symposium Systems in Aviation. British Airline Pilots Association, 1971.

41. Gordon-Johnson, P. *The Development of Noise Abatement Approach Procedures and a Pilot Subjective Assessment of Workload During the Approach and Landing Phase of Flight.* Unpublished Royal Aircraft Establishment Report, 1975.

42. Roscoe, A.H. *Use of Pilot Heart Rate Measurement in Flight Evaluation.* Aviat. Space Environ. Med. 47, pp.86-90, 1976.

43. Cooper, C.E. *The Pilot-aircraft Interface.* Symp. Vehicle Technology for Civil Aviation, Seventies and Beyond. NASA, Washington DC, 1971.

44. Nicholson, A.N. et al. *Influence of Workload on the Neurological State of a Pilot During the Approach and Landing.* Aerospace Med. 44, pp.146-152, 1973.

45. Wallick, S.L. *Internal Environment and Flight Deck Layout.* Symposium. Flight Deck Environment and Pilot Workload. Royal Aeronautical Society, London, 1973.

46. ter Braak, F. *High Workload Tasks of Aircrew in the Tactical Strike, Attack and Reconnaissance Roles.* In: Conference Proceedings No.146. Simulation and Study of High Workload Operations. AGARD, Paris, 1974.

47. Brown, I.D. *The Measurement of Perceptual Load and Reserve Capacity.* Trans: Ass. Indust. Med. Offrs. 14, pp.44-49, 1964.

48. McRuer, D.T. *Neurological Applications of Man-machine Systems Analysis.* In: Technology and the Neurologically Handicapped. NASA SP – 363, pp.191-196, Washington DC, 1971.

49. Wanner, J-C. *Piloting and Flying Qualities of the Next Generation of Aircraft.* Aeronautical J. 77, pp.593-605, 1973.

50. McRuer, D.T. Kendel, E.S. *Mathematical Models of Human Pilot Behaviour.* AGARD AG-188, Paris, 1974.

51. Curry, R.E. Ephrath, A.R. — *Monitoring and Control of Unreliable Systems.* In: Conference Preprints. Monitoring Behaviour and Supervisory Control. pp.134-143, Scientific Affairs Div NATO Brussels, 1976.

52. Storm, W.F. et al. — *The Effects of Two Stressors on Traditional and Engineering Analogs of Cognitive Functioning.* In: Conference Proceedings No.181. Higher Mental Functioning in Operational Environments. AGARD, Paris, 1975.

53. Wanner, J-c. — *Présentation des Information Nécessaires pour le Décollage et l'Atterrissage.* In: Conference Proceedings No.160. Take-off and Landing. AGARD, Paris, 1974.

54. Rault, A. — *Pilot Workload Analysis.* In: Conference Preprints. Monitoring Behaviour and Supervisory Control. pp.350-364. Scientific Affairs Div. NATO, Brussels, 1976.

55. Christensen, J.M. — *Human Factors Engineering Consideration in Systems Development.* In: Measurement of Man at Work. Singleton, W.T, Fox, J.G and Whitfield, D. (Eds). Taylor and Francis, London, 1971.

56. Sheridan, T.B. — *Preview of Models of the Human Monitor/Supervisor.* In Conference Proceedings Preprints II. Monitoring Behaviour and Supervisory Control. pp.11-18, Scientific Affairs Div. NATO, Brussels, 1976.

CHAPTER 2

SUBJECTIVE ASSESSMENT PILOT OPINION MEASURES

by

Flight Lieutenant George A.Ellis
Flight Systems Squadron
Royal Aircraft Establishment
Bedford, England MK41 6AE

CONTENTS

SUBJECTIVE ASSESSMENT PILOT OPINION MEASURES

2.1 INTRODUCTION

Pilot opinion has traditionally played an important part in the assessment of workload. For instance, Gerathewohl[1] wrote:- "Subjective pilot rating is still the common method of assessing the handling qualities of an aircraft and the total workload in determining its suitability for an intended mission". It is likely that this will continue to be the case for the foreseeable future.

In the more academic studies of workload, and where workload has been measured in isolation from other factors such as handling qualitities, it has been normal to use subjective assessments as a back up for other measures. The chosen measurements have been placed in comparison with pilot opinion; depending on the researcher, the results have been used to comment either on the accuracy of the scientific measurement or on the accuracy of the subjective assessment. An example of the latter approach can be taken from Krzanowski and Nicholson[2] who observed that: "The correlation between subjective measures and physiological changes suggested that workload assessments by the pilot may be of value".

There is, however, an increasing feeling that greater emphasis should be placed on subjective measurement, especially in cases where either the task is too complicated for an objective measuring technique, or extrapolation is required to other flight tasks. In making his assessment, the pilot has the advantage of letting his experience and feelings influence his judgement, and he can take into account any factor that he considers relevant. However, these feelings might be construed as prejudices and he must be able to uphold, explain and defend his assessment in a clear and logical fashion; this can be difficult when hard data is lacking and when his judgement is embarrasing to others. Nevertheless, where the opinion of pilots, and especially trained test pilots, points clearly in a particular direction, that opinion should carry the largest weight. As Gartner and Murphy[3] have pointed out: "When experiential conceptualizations of workload are accepted, the pilot's direct perception or estimation of his feelings, exertion, or conditions may provide the most sensitive and reliable indicators".

Of course, a pilot's subjective assessment suffers the disadvantage of not being objective; it is difficult to analyse, and it cannot readily be quantified. Nevertheless, experience with handling qualities assessments has shown that pilot opinion, properly expressed within the framework of a rating scale, can provide a valid scientific measure.

Considering the influence that pilot opinion ought to have, it is disappointing that the subjective assessment of workload has not been studied with the thoroughness applied to the other methods. Researchers who have used subjective ratings to compare against their scientific results have devised rating scales and questionnaires, often very good ones, and used the results. Since the subjective assessments were not the main purpose of the experiment, there has seldom been an attempt to comment on the validity and usefulness of the chosen subjective assessment technique, or to make recommendations for its future improvement.

There is now a definite need for a standardized approach to the problem, so that a widely acceptable method for subjective assessment can be developed and adopted. Even if it cannot be agreed that the method is optimum, standardization on a single method should bring considerable benefits. Pilot assessment of workload could then be properly influential, not only as a measure to back up other methods, but also as a primary measure in its own right.

The aim of this chapter is to briefly review concepts of workload and methods of subjective assessment, and to discuss them from the point-of-view of the test pilot. By doing this, it is hoped that researchers of all disciplines will be helped in their understanding of workload and will appreciate the large contribution that pilots can make to such experiments. The chapter is mainly concerned with the kind of day-to-day evaluation familiar to most test pilots, both in flight and in ground-based simulators, rather than with elaborate academic experiments. The author is a practising experimental test pilot and so most of the comments which follow are based on his own experience, and that of his colleagues at Bedford, of making subjective evaluations of workload and handling qualities (mainly the latter) rather than on an extensive study of the literature.

2.2 CLASSIFICATION AND DEFINITION OF WORKLOAD

2.2.1 Short-term and Long-term Workloads

It is useful to classify workloads according to the lengths of time for which they are being considered. The time-scales defined by Howitt[4] and Benson[5] have already been mentioned in Chapter 1. The pilot would normally only concern himself with assessing short-term workloads – the immediate workloads of a flight phase or sub-phase, or possibly the combined immediate workloads of a series of phases or a whole flight. He would not assess the duty-day or long-term workloads, though he would probably comment on them if he felt that they would be affected by the vehicle or tasks whose short-term workloads he was assessing.

This chapter is, therefore, concerned with the assessment of the immediate workload – the workload experienced over any particular short period of time.

2.2.2 Task-related and Pilot-related Workloads

Gartner and Murphy[3] noticed that the definitions of workload adopted by researchers could be gathered into three groups (see also Chapter 1). The first one is workload as a set of tasks demands. This approach is concerned with what is required or demanded of the crew in the performance of a task, but it does not measure the resulting response of the crew. Secondly, there is workload as operator effort. This concept looks at how hard the pilot is working – the amount of effort and attention he is giving to the task. Finally, there is workload as activity or accomplishment – the actual task performance or the products of pilot activity.

The first two sets of ideas, task-related and pilot-related, are the ones more usually used when defining workload. Unfortunately, these two approaches are not synonymous; one of them must be chosen as the basis for the definition of workload that will be used in the subsequent assessment. The choice is crucial, for, as Thorne[6] has said: "I doubt if we shall ever be able to measure task difficulty and operator capacity on the same scale".

A pilot-related definition of workload is to be preferred for the purposes of subjective assessment, for the following reasons:

(a) The assessment should attempt to measure the way that workload affects the pilot, and such a measurement will only result from a consideration of how hard the pilot is working. This point is also covered in Chapter 1.

(b) Implicit in the idea of pilot effort is the concept of rate of work. The pilot feels that his workload is higher if he has to compress his actions and decisions into a shorter timescale, even if those actions and decisions remain the same.

(c) A recent survey among civilian air transport pilots in Britain[7] showed a substantial preference for thinking of workload as pilot-related rather than task-related, even though many of the subjects felt that task demands remained an important consideration.

(d) The concept is already used in the most widely accepted method for subjective assessment of aircraft handling qualities, the Cooper-Harper rating scale. Cooper and Harper[8] defined workload as "the integrated physical and mental effort required to perform a specified piloting task". The definition is a good one, test pilots are familiar with it, and it is sensible to standardize on it for all subjective evaluations.

Of course, considerations of task demands and task performance remain relevant. It may be useful to differentiate between predicted workloads, based on task demands, and actual (or measured) workloads based on pilot effort. Performance is important because it represents the end result of the pilot's efforts; the effects of workload can often be put properly into context by relating levels of workload to levels of performance.

2.2.3 Physical and Mental Workloads

The Cooper-Harper definition of workload encompasses physical and mental effort. Physical workload is defined as "The effort expended by the pilot in moving or imposing forces on the controls during a specified piloting task". Mental workload is not defined, but is left to pilot evaluation or assessment by indirect methods. Mental workload includes such tasks as perception, information processing and decision making.

Some researchers[9] have included a third category, perceptual workload. Although the distinction may well be significant when workload is related to task demands, the inclusion of this separate category within a subjective assessment should be resisted for the following reasons. First, perception is a mental task and there is no good reason to single it out in subjective assessments. Secondly, the pilot should be faced with questions and choices that are as simple as possible; if necessary he can probably divide his workload fairly easily into its physical and mental parts, but any further subdivision should not be demanded of him. Finally, if any principle contributions to mental workload are obvious to the pilot, he will make mention of them in his qualitative comments.

2.3 PRINCIPLES AND METHODS OF SUBJECTIVE ASSESSMENT

2.3.1 The Relationship of Workload and Handling Qualities

In this discussion of methods of subjective assessment, frequent mention will be made of handling qualities assessments. The reasons for this fall into two main groups.

Test pilots are well acquainted with the techniques employed in the subjective rating of handling qualities, and the consideration of workload is an essential step in the rating process. The judgements of workload made in these circumstances are unlikely to be formerly expressed and may well be instinctive, but a close relationship exists between the subjective assessments of handling qualities and pilot workload.

Most of the work that has been done in developing and evaluating methods of subjective assessment has been in the field of handling qualities. Much, therefore, can be borrowed from this work and read across to the study of workload. Very importantly, the subjective assessment of handling qualities had developed to the extent that there is a widely accepted handling qualities rating scale, the Cooper-Harper scale; this scale is discussed in paragraph 2.3.4.

2.3.2 Spare Capacity

One of the most useful aids to the measurement of workload has been the idea of spare capacity. A man's capacity for work is finite and unless he is working at his limit he must be able to increase his workload to some extent. Therefore it should be possible to quantify his workload by measuring the amount by which he can increase his effort: in other words his spare capacity.

A common method used for the objective measurement of spare capacity is to give the pilot a secondary task and to score his performance in it, whilst he continues with the primary task[10,11]. Secondary task techniques are reviewed in Chapter 4. It is worth noting, though, that there are several drawbacks to using secondary tasks. In addition pilots would probably resent their presence and may well find that they interfere with the assessment and so they are not really suitable for use within a technique for subjective assessment.

McDonnell[12] performed an elegant experiment that not only overcame some of the disadvantages of using secondary tasks, but also compared the method with subjective assessment. The difficulty of the secondary task was made to vary with the pilot's performance in the primary task, and the secondary task scores were compared with Cooper ratings that the subjects had given to the primary task alone. The results showed a very good correlation and have been quoted[13,14] as evidence that subjective ratings can be used as measures of spare capacity and workload.

It was found by Ellis and Roscoe[7] that pilots are in favour of thinking about workload in terms of spare capacity. Therefore, a subjective measurement of workload linked with the notion of spare capacity would seem to be worth pursuing: it is likely to be dependable and readily acceptable to pilots.

2.3.3 Rating Scales

The use of rating scales results in the allocation of a numerical value to the quantity that is being measured. Not unnaturally, researchers wish to use statistical and mathematical processes on the numbers so obtained, and so most of the rating scales that have been devised have been intended to be linear.

One common technique, used for instance by Nicholson[15] and Rolfe[9] is the 10cm line method: the pilot is asked to indicate his opinion by making a mark on a line whose ends are labelled with the opposite extremes of opinion (e.g. Extremely Difficult and No Difficulty); the rating is then taken from the position of the pilot's mark. The 10cm line method has several disadvantages (and these are shared by many other rating scales). It is by no means certain that one pilot's linear scale will be linear against another pilot's; this is likely to be important when small sample sizes are used. Secondly, not all researchers have managed to make the ends of their scales reflect true opposites[16]. Thirdly, there is a natural tendency for pilots to commence rating at the middle of the scale to allow for movement either way[16]. Finally, perhaps the most important drawback of the technique is the tendency to ascribe to it an unwarranted degree of fineness. It may well be possible to measure the pilot's marks to the nearest millimetre and then to analyse the results, but to what extent is this valid.? Krzanowski and Nicholson[2] said: "The continuous line technique for subjective assessment may give an unwarranted impression of accuracy and the question arises whether a box technique would be more appropriate. This may indicate a greater significance of the movements of the assessment and reduce the variance of assessment in high workload situations".

Another method for trying to get a linear measure is to ask the pilot to state a numerical rating on a scale of, typically, 7, 9 or 10 points. Often, the subject is guided to a rating by the allocation of adjectives to certain parts of the scale. A rating scale of this type was described by Borg[17] in which there were 15 values, the odd values being anchored with the aid of verbal expressions (the aim of Borg's experiment was to correlate the rating scale and the physical work level in a non-aviation physical task).

Although rating scales have shown good correlation with objective measurements in purely physical tasks, there is no reason to expect that any of these scales should be linear with respect to any physical variable when mental effort is also included. However, if linearity is required of a scale, it should aim to be linear in a way that is subjectively acceptable. McDonnell[12] went to some lengths to establish an underlying psychological scale, and based on this he proposed a 7-point scale for handling qualities. Adjectives describing the favourability of the qualities were given positions on the scale, which is reproduced in Figure 1.

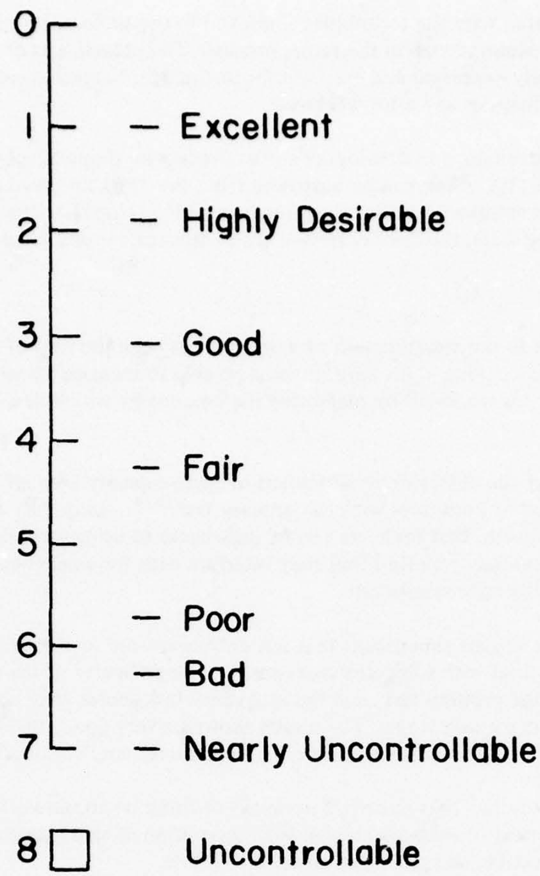

Fig. 1 Favorability of handling qualities

Rating scales have also been used that have been accepted as being non-linear; researchers have often taken means and standard deviations of this type of rating, though caution should be exercised when subjecting results like these to any analytical process. The most important scale in this category is the Cooper-Harper scale for aircraft handling qualities (paragraph 2.3.4). Spyker and his colleagues[18] decided to use elements of the Cooper-Harper scale to get subjective ratings on their workload experiments. The subjects were presented with a series of six questions and were asked to indicate an answer to each one by choosing one of a limited number (5 to 9) of phrases describing opinions; each answer was allocated a numerical value that corresponded to the position it would have on a scale of the Cooper-Harper type. Two of Spyker's sets of questions and answers are shown in Figure 2.

2.3.4 The Cooper-Harper Rating Scale

The Cooper-Harper rating scale for handling qualities is such an important scheme for subjective assessment that it deserves separate mention. It was very carefully developed, test pilots are used to using it, and it is widely accepted as the standard scale.

Cooper and Harper had a very clear and logical approach to the problems of subjective assessment, and their report[8] is well worth studying. Their scale is reproduced in Figure 3, and the following points should be noted about the scale and the methods of using it:

(a) The scale is more than one of pure comparison. Whereas a pilot can be expected to place a number of vehicles (or configurations) in order of desirability, his Cooper-Harper rating for any of them is intended to be repeated whatever the qualities of the other vehicles under assessment. Thus, if one example is given a rating of 4 in an experiment where all others lie between 6 and 8, it should also be rated 4 if its rivals were to lie between 1 and 3.

(b) Despite this, the scale is essentially a comparative one and so does not present the pilot with an unreasonably difficult task. McDonnell[12] commented: "Rating scales are subjective in nature and therefore are scales of comparison". The 'absolute' value-judgements that pilots are expected to make are based on their own empirical knowledge; the success with which these judgements can be made is a consequence of the careful definition of each value ('satisfactory' etc) and on the wisdom and experience of the assessing pilots.

(c) The pilot is drawn towards the eventual rating through a step-by-step process. The value judgements that he makes are presented as a series of decisions. The dichotomous choices at each stage of the decision 'tree' are

II. In my opinion control over the simulated aircraft was:

- ☐ Extremely easy to control with excellent precision (0.5)
- ☐ Very easy to control with good precision (2.5)
- ☐ Easy to control with fair precision (4.5)
- ☐ Controllable with somewhat inadequate precision (6.5)
- ☐ Controllable, but only very unprecisely (7.5)
- ☐ Difficult to control (8.0)
- ☐ Very difficult to control (8.5)
- ☐ Nearly uncontrollable (9.0)
- ☐ Uncontrollable (10.0)

III. In my opinion the demands placed on me as the pilot were:

- ☐ Completely undemanding, very relaxed and comfortable (2.5)
- ☐ Largely undemanding, relaxed (3.5)
- ☐ Mildly demanding of pilot attention, skill, or effort. (5.5)
- ☐ Demanding of pilot attention skill or effort (6.5)
- ☐ Very demanding of pilot attention, skill, or effort (7.5)
- ☐ Completely demanding of pilot attention, skill, or effort (8.5)
- ☐ Nearly uncontrollable (9.0)
- ☐ Uncontrollable (10.0)

Figure 2

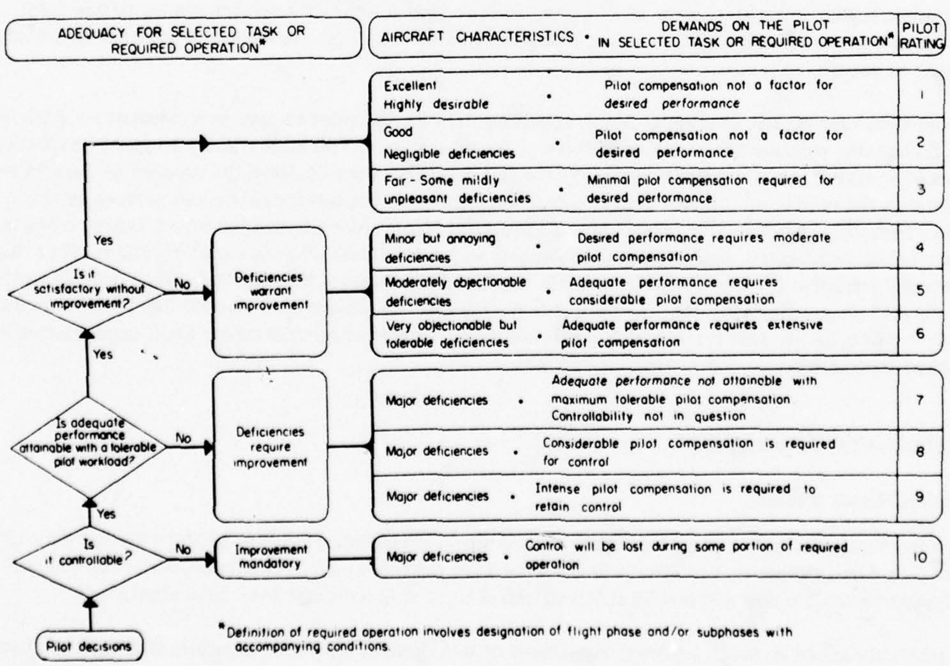

Figure 3

fairly simple, and once the vehicle under assessment has been placed within the value 'boundaries' (3½, 6½, 9½) the pilot has to choose one of only three values. Although he was unhappy about the nature of the boundaries, McDonnell[12] acknowledged the value of the decision tree as an aid to assessment.

(d) The scale is aimed towards the practical application of the vehicle under assessment. The pilot's judgements are all made in the context of the defined task or mission. The pilot is therefore asked to consider what can reasonably be expected of man and machine in the circumstances rather than to assimilate some hypothetical set of values applicable to all conditions. This property of the scale means that the task or mission must be clearly defined, and in a way that is acceptable to the assessing pilots; consequently, the ratings are valueless unless the definitions (and instructions to the pilots) are quoted alongside the results.

(e) The Cooper-Harper rating does not provide a complete assessment. It gives a shorthand guide to the worth of the vehicle, but the pilot should also state why he arrived at the rating and what improvements he thinks are necessary.

(f) The scale is very practicable. One learned, it is easy to use and so it is suitable not only for laboratory experiments but also for real flight conditions. A pilot can give a rating and make a few cryptic comments while he is flying an aeroplane, a circumstance in which he cannot be expected to go through an assessment ritual that is long and complicated.

(g) The Cooper-Harper scale uses workload in a very specific but limited manner. Workload is always related to the task; overall workload is judged against a standard of tolerability ("Is adequate performance attainable with a tolerable pilot workload?" – from Figure 3); other workload decisions are based on the concept of compensation (compensation is defined as "The measure of additional pilot effort and attention required to maintain a given level of performance in the face of deficient vehicle characteristics").

(h) The scale is ordinal. Naturally enough, researchers would prefer the scale to be a linear or interval one, and so have criticized it. Nevertheless, the construction of a practical scale of demonstrated linearity has not yet been achieved, and the many advantages of the Cooper-Harper outweigh its disadvantages. Some researchers take means and standard deviations of Cooper-Harper ratings, and although it might be convenient to express results in this way, caution should always be exercised when manipulating the numbers derived from this scale. For example, an average rating from a number of pilots might obscure the fact that one of them gave a much lower (or higher) rating than the others. The reasons for this isolated result may be simple, but should not be ignored.

2.3.5 Questionnaires and Pilot Reports

Subjective assessments and questionnaires are inseparable. The pilot is faced with a questionnaire, whether it is in the form of a few simple questions from the researcher or a detailed multi-question document. Questions fall into two groups. First, there are the forced-choice questions, to which the pilot can make only a limited number of replies: at their simplest these will be dichotomous (was there less workload on run A or run B?) but rating scales are more typical examples of this type of question. Secondly, there are open-ended questions, in which the pilot is asked to comment on some aspect of the experiment; the pilot report, in which comments are made without guidance from the researcher, can be considered analogous to the reply to an open-ended questionnaire, although it is an extreme example of the genre. Open-ended questions produce replies that can be very awkward to analyse, and so there is some reluctance to ask them. However, it is important to know why a pilot has given a rating or a particular forced-choice answer, and so some unforced pilot comment must be sought.

Questionnaires vary greatly in length. Some experimenters ask the pilot to rate, or comment on, perhaps fifty[19] aspects of a flight task, whereas others ask only three or four questions. The pilot will try to give honest answers to all the questions he is asked, but it is quickly learnt by the test pilot that he must limit the amount of data he tries to gather from any one experimental run; if he attempts to observe and remember too much, he will achieve less than he might otherwise do. Therefore, whereas it is reasonable to ask a pilot to consider a large number of factors when long-term workloads are being investigated, or for specific laboratory experiments on pilot opinion[9,19], the number of questions should be severely curtailed for the sort of immediate workload assessments that the test pilot normally undertakes. An example of the size of questionnaire that the author feels is about the maximum he would like to face comes from Schultz et al[20]: pilots had to rate the overall aircraft and three aspects of control (using the Cooper-Harper scale) and they were asked to comment on eight subsidiary factors.

2.4 PRACTICAL CONSIDERATIONS

2.4.1 Aims and Expectations

It is unfortunately the case that not all researchers define with sufficient clarity or accuracy the aims of their experiments. Likewise, people do not always know what to expect from subjective assessment. As with any other form of scientific experiment, success will not be achieved unless these shortcomings have been eliminated.

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. This is not always the case, partly because of the loose employment of the term "workload". It is not unknown for researchers to ask for a workload assessment as a primary measure in exercises whose real aim has been the improvement or assessment of an aircraft or aircraft system; in these cases, the most appropriate subjective consideration is one of handling qualities and the assessment should be based on the Cooper-Harper scale. 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.

The researcher's expectations should be realistic. Experience indicates that subjective assessment is more likely to provide valid answers when the limitations of the technique are understood and allowed for. The results will be qualitative and unsuitable for detailed mathematical analysis, the pilot should be asked to consider only a few factors at a time, and the best results will be achieved from a well-structured rating scale backed by pilots' comments.

2.4.2 Design of a Rating Scale

If a rating scale for workload is to be successful, it is likely that it will have been constructed along lines similar to those of the Cooper-Harper scale. Of course, the Cooper-Harper scale has been used in connection with workload measurement[12], and so the question arises whether it is an adequate scale for workload rating. The answer was provided by Geratewohl[1]: "Although workload is seen as inextricably tied to the assessment of such characteristics as compensatory system monitoring and precision of control, judgements of perceptual or mental effort involved in this process are generally not obtained. Hence, subjective pilot ratings of handling qualities, as accurate as they may be in regard to control desirability or difficulty, do not contribute to workload determinations, since they are only loosely connected to task demands and pilot response".

The most straightforward scale to construct would be one that parallels the Cooper-Harper as closely as possible. The decisions would all have to be taken in the context of the task or mission under consideration; the same categories of Satisfactory, Unsatisfactory and Unacceptable could be reached by a decision tree; the difference would be that pilot effort would be the criterion, and the assessment would probably be aided by asking the pilot to consider his spare capacity. A scale of this type is likely to be practicable in cases where workloads within the same type of task are being compared. Because the scale would be equivalent to the Cooper-Harper, it is to be expected that the workload ratings and the C-H ratings would be the same in many cases; differences are, however, likely to occur. To give some examples, if a pilot does a series of runs at the same task in the same vehicle, his workload will probably go down as he becomes accustomed to the conditions of the day; additionally, he might raise his workload in order to improve his performance, or he might lower his workload knowing that his performance will remain acceptable.

A task-related scale of the type described in the last paragraph might well be very practicable, but it would have only limited applications. It would not allow comparisons to be made of workloads in different tasks. How much harder does a man work when he is landing than when he is in cruising flight? How does the workload during a bombing attack compare with the workloads just mentioned? Any subjective scale that was expected to show agreement with an objective or physiological measure would have to be able to answer questions like those just posed. The crux of the matter is that experienced pilots are likely to be able to judge whether a workload is satisfactory for the flight phase (or sub-phase) under consideration, but a task such as landing demands a build-up to a high degree of accuracy over a short time period, and a workload satisfactory for landing is likely to be higher than that tolerable in cruising flight. Is the pilot able to quantify this on an "absolute" workload scale?

If a scale is to be constructed that leads to the allocation of an "absolute" workload rating for all conditions, it is, once again, likely to be practicable only if it is designed according to well-proven principles. The pilot should be guided to his rating by a dichotomous decision tree that leads to workload descriptors. The rating should reflect pilot effort, and to help the pilot describe his level of effort, the framework of the scale should enable him to consider the length of time for which he could (or would wish to) sustain that effort, and the extent of his spare capacity. The scale of "absolute" workload is unlikely to be suitable for all applications; the task-related scale would be a more sensitive measure and so would be of more use during the majority of evaluations. It remains that, unlike the handling qualities case, more than one workload scale is likely to be necessary.

2.4.3 The Choice of Subject Pilots

For any experiment involving human subjects, and especially for an experiment in which subjective assessments are being made, it is vital to ensure that the subject pilots will give valid results. The point may seem to be an extremely elementary one, but all too often reports are published whose results and conclusions must be treated with suspicion because of the employment of unsuitable subjects (or an insufficient number of subjects) in the experiments.

Pilots are not representative of the community at large when it comes to controlling aircraft or aircraft simulators. The training of a pilot is a long and expensive undertaking, in the course of which his judgement of airborne circumstances is formed and refined, and his reactions in different situations become more consistent and safe; in other words he must acquire a level of airmanship that needs not only to be taught but also to be developed by exposure to the practical problems of flight. Unsuitable individuals are discarded at each stage, and some pilots who have been trained to very advanced levels never become truly proficient aviators and have to be returned to jobs on the ground. However good and thorough any training system has become, the pilot is not considered to be an experienced and expert airman until he has completed several years of productive flying, in the course of which his judgement has been allowed to mature. Except in specific, exceptional and specialized experiments for which inexperienced pilots are required, subjects should only be chosen who are mature and experienced pilots.

Just as piloting skills are not acquired by every individual, so the skills of making subjective assessments are not acquired by every experienced pilot, and these assessment skills must be developed in the individuals who are required to take part in aeronautical research. The point has been well made by Schultz[20]. "To seek a relationship between subjective pilot ratings and system performance that can be applied to real-world situations, it is clearly necessary that experienced handling qualities evaluation pilots be used as subjects in the experiment".

Trained test pilots are likely to be the most suitable subjects. Given the chance, other pilots may well develop into skilled assessors, but more dependable (and more widely accepted) results will be obtained by the employment of test pilots. Test pilots have been carefully chosen as being suitable for the job, they have a wide experience of different types of aircraft, they have been specially trained in the art of making assessments, and the nature of their employment is such that they are in good practice at looking critically at aircraft systems, and in making subjective evaluations. In addition the experienced test pilot will have taken part in many experiments both in the air and in the laboratory, and so he will be able to help the researcher to set up a good experiment.

2.4.4 Simulation

There is an element of simulation in many full-scale assessments: it would be most unusual, for instance, to assess a new warplane by sending it straight into war, and many types do not see action (if they see it at all) until they have been in service for some years; yet they all have to be assessed for their war roles. However, by far the most common simulation problems are those posed by the use of ground-based research simulators, If any degree of simulation is present in an experiment, the results will have to be extrapolated to the real-life situation. Clearly, the further removed the experiment is from reality the more tentative must be the extrapolation. Two of the references are now quoted. Schultz[20], in connection with his simulator experiment, said: "For such an experiment a full-task situation must be presented to the pilot or he will feel that he is involved in a game that, no matter how interesting, is not related to flying an airplane. Therefore, he cannot be expected to perform as a pilot would in a real situation". Cooper and Harper[8] discussed several aspects of simulation, and wrote: "Previous studies have shown that sophistication is not necessarily the key to simulator usefulness although it can extend the range of application. Deciding "what a pilot rating applies to" (specific task or flight phase), and the completeness of the simulation will determine the degree to which pilot extrapolation is to be relied on. Neither the pilot nor the engineer retains confidence in the results if the need for extrapolation of observed results becomes too great . . . It is felt that careful planning and agreement on program objectives, mission definition what is being rated and the execution of the experiment can limit the uncertainties of extrapolation".

2.4.5 Subdivision of the Task

A form of task sub-division frequently used is to ask the pilot to consider separately the different concurrent parts of his task. Schultz[20] asked the pilot to rate the longitudinal mode only, the lateral-directional mode only, the total overall airplane, and whether or not the airplane could be landed. Nicholson and his colleagues[2,15] asked the pilot to consider, and mark, some five factors in addition to overall difficulty, namely aircraft, navigational aids, meteorological conditions, physical features of the airport and control procedures. In both these cases, the pilots were successful in making multi-ratings, and so the technique would seem to be practicable. Nicholson et al were dissatisfied with the overall ratings at high workload levels, because they were at variance with the other data collected. However, researchers should be very wary of rejecting any pilot opinion. The well-accepted hypothesis that there is an inverse-U relationship between performance and arousal state discourages any idea that workload, especially at high levels, will be additive. The pilots' opinion must be that the overall rating is the important one, and that irregularities are due to the non-linearity of man's behaviour at high workload.

2.5 CONCLUDING REMARKS

Some readers may feel that the approach to this chapter has been too elementary, and that many of the points raised have been unnecessarily basic. It is felt that this approach is justified by the lack of understanding, in many quarters, of the value and real meaning of pilot opinion, and by the fact that research involving pilot opinion is not yet free from unsound experimental techniques. Workload research is a good example of a field that is best served by a multi-disciplinary approach. In order that pilots' subjective assessments can properly be made and utilized, the main conclusions propounded in the chapter are here repeated:

(a) Workload should be clearly defined, and the definitions should be operator-related rather than task-related. The most suitable definition comes from Cooper and Harper[8]: "The integrated physical and mental effort required to perform a specified piloting task".

(b) The best way to express a subjective assessment is through a simple rating scale amplified by pilots' explanatory comments.

(c) Pilot ratings are qualitative, and attempts to subject them to inappropriate forms of numerical analysis should be resisted.

(d) Any rating scale should be designed using the principles employed in the Cooper-Harper scale[8].

(e) Pilots should not be overburdened, and should be asked to answer only a strictly limited number of questions in any assessment.

(f) Subject pilots should be carefully chosen. The best results will be obtained by using experienced evaluation test pilots.

(g) Great care should be taken when extrapolating the results of simulator exercises.

REFERENCES

1.	Gerathewohl, S.J.	*Definition and Measurement of Perceptual and Mental Workload in Aircrew and Operators of Air Force Weapon Systems:* A Status Report. In AGARD CP-181: Higher Mental Functioning in Operational Environments. Paris, April 1976.
2.	Krzanowski, W.J. Nicholson, A.N.	*Analysis of Pilot Assessment of Workload.* Aerospace Medicine. 43, pp.993-997, 1972.
3.	Gartner, B. Murphy Miles, R.	*Pilot Workload and Fatigue:* A Critical Survey of Concepts and Assessment Techniques. NASA TN D-8365. Washington DC, November 1976.
4.	Howitt, J.S.	*Flight-deck Workload Studies in Civil Transport Aircraft.* In AGARD CP-56: Measurement of Aircrew Performance. Paris, December 1969.
5.	Benson, A.J.	*Technical Evaluation in AGARD CP-56:* Measurement of Aircrew Performance. Paris, December 1969.
6.	Thorne, R.G.	*Pilot Workload – A Conceptual Model.* In AGARD CP-119: Stability and Control. Paris, November 1972.
7.	Ellis, G.A. Roscoe, A.H.	*Pilot Workload:* A Survey. (In Preparation).
8.	Cooper, E. Harper, P.Jr.	*The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities.* AGARD Report 567. Paris, April 1969.
9.	Rolfe, J.M. et al.	*Evaluating Measures of Workload Using a Flight Simulator.* In AGARD CP-146: Simulation and Study of High Workload Operations. Paris, October 1974.
10.	Rolfe, J.M.	*The Secondary Task as a Measure of Mental Load.* In: Measurement of Man at Work. Singleton, W.T, Fox, J.G. and Whitfield, D. (Eds). London, 1971.
11.	Corkindale, K.G.G.	*A Flight Simulator Study of Missile Control Performance as a Function of Concurrent Workload.* In AGARD CP-146: Simulation and Study of High Workload Operations. Paris, October 1974.
12.	McDonnell, J.D.	*Pilot Rating Techniques for the Estimation and Evaluation of Handling Qualities.* AFFDL-TR-68-76. Wright-Patterson AFB, Ohio, December 1968.
13.	Ashkenas, I.L.	*Lead Discussion on Hosman, R J A W. Pilot Vehicle Analysis.* In AGARD CP-106: Handling Qualities Criteria. Paris, June 1972.
14.	McRuer, D.T. Krendel, E.S.L.	*Mathematical Models of Human Pilot Behaviour.* AGARD-AG-188. Paris, January 1974.
15.	Nicholson, A.N. Hill, L.E. Borland, R.G. Ferres, Helen M.	*Activity of the Nervous System During the Let Down, Approach and Landing.* A Study of Short Duration High Workload. Aerospace Medicine, 41, pp.436-446, 1970.
16.	Allnut, M.F. Bolton, C.B.	*The Construction and Usage of Questionnaires.* Royal Air Force Institute of Aviation Medicine S120. Farnborough, Hants, June 1974.
17.	Borg, G.	*Psychological and Physiological Studies of Physical Work.* In: Measurement of Man at Work. Singleton, W.T, Fox, J.G. and Whitfield, D. (Eds), London, 1971.

18. Spyker, D.A. Stackhouse, S.P. Khalafalla, A.S. Mc Lane, R.C. *Development of Techniques for Measuring Pilot Workload.* NASA CR-1888. Washington DC, November 1971.

19. Lindsay, S.J.E. Evans, R.L. *A Semantic Analysis of Workload.* Royal Air Force Institute of Aviation Medecine S118. Farnborough, Hants, November 1974.

20. Schultz, W.C. Newell, F.D. Whitbeck, R.F. *A Study of Relationships Between Aircraft System Performance and Pilot Ratings.* NASA CR-1643. Washington DC, July 1970.

CHAPTER 3

PHYSIOLOGICAL METHODS

by

Alan H.Roscoe
Medical Department
Royal Aircraft Establishment
Bedford, England MK41 6AE

CONTENTS

PHYSIOLOGICAL METHODS

3.1 INTRODUCTION

In 1966, Westbrook and his co-authors[1] concluded a paper on handling qualities and pilot workload by stating: "There is no question but that the handling qualities engineer should broaden his idea of workload and make a concentrated effort to apply the ideas and measurement tools of the physiologist and the psychologist to the quantitative measurement of workload". Physiological, or psycho-physiological techniques are discussed in this chapter with a view to using them to assess levels of pilot workload. Only one of the broad concepts of workload discussed in Chapter I is appropriate to the use of physiological techniques for its assessment, namely: the physical and mental effort required by the pilot to meet the demands of the flight task.

For some time, well established and reliable physiological techniques have been used to calculate levels of physical workload in terms of energy expenditure. For example, estimating oxygen consumption, measured directly by using some form of gas collecting device or indirectly by measuring heart rate, is a common practice of the work physiologist. However, pilot workload during normal flight contains a relatively small amount of physical load, being classed as sedentary or light work.[2,3] On the other hand, the non-physical content of pilot workload may be quite high. Unfortunately, estimating energy expenditure using conventional methods does not give anything like a true picture of the total workload involved in performing a complicated flight task such as a take-off or an approach and landing. We have therefore to consider the application of physiological methods to the study of mental as opposed to physical effort.

In 1921 Golla[4] discussed the fundamental mechanisms of cerebral activity and pointed out that mental as well as physical effort causes changes in physiological activity. Physiological methods for assessing mental activity have been developed over many years from techniques originally evolved for phychophysiological research into such aspects of human behaviour as response to drugs, vigilance, fatigue and into problems associated with neuropsychiatric illness. In this chapter, we shall review the various methods and their possible application to workload in the context of the flight task.

3.2 THE CONCEPT OF AROUSAL

The rationale of using physiological measures to measure mental load is based on the concept of "activation" or "arousal", a state of preparedness of the body associated with increased activity in the nervous system. Duffy[5] described: . . . "the level of activation of the organism as the extent of release of potential energy, stored in the tissues of the organism as this is shown in activity or response". She also suggested that: ". . . it would be possible to define activation as the arousal which occurs in the absence of physical exertion". It has been suggested by Welford[6] that any task demanding an effort or ". . . which is in some way challenging", raises the level of arousal. Arousal may be considered as a continuum with sleep or unconciousness at one end and hyperexcitation or extreme agitation at the other.

Several investigators have studied the relationship between activation or arousal and performance. Duffy[5] concluded from experimental evidence that ". . . the degree of activation of the individual appears to affect the speed, intensity and coordination of response and thus to affect the quality of performance". She also observed that in general, the optimum level of activation appears to be a moderate level, with the curve expressing the relationship between performance and activation taking the form of an inverted 'U'. Other authors have argued this relationship, although there is only meagre experimental evidence to support it. In 1908, Yerkes and Dodson[7] described an inverted 'U' shaped relationship between motivation and learning and recently Davey[8] has shown a similar relationship between arousal and physical exercise. Welford[9] proposed an inverted 'U' hypothesis as a model to describe the relationship between arousal resulting from stress, and performance.

The concept of arousal, which is now accepted by most authors as being synonomous with activation, is a convenient way of relating physiological activity to pilot workload. It can be argued that as the flight task, the effort put out by the pilot and the resulting performance are related, and that as arousal and performance are also related, then levels of physiological activity should provide realistic estimates of workload levels. Implicit in this argument is the need to monitor performance whenever physiological activity is measured for the purpose of assessing workload.

3.3 PHYSIOLOGY

3.3.1 Control Mechanisms

A brief account of the physiological control mechanisms associated with arousal and which also regulate systems suitable for measurement is given in this section.

The nervous system can be divided into two main components, the central nervous system (CNS) which is made up of the brain, brain stem and spinal cord, and the peripheral nervous system. The peripheral nervous system can be split into the somatic division which conducts impulses to and from the various voluntary muscles and sensory organs, and the autonomic or involuntary nervous system (ANS).

This latter division is of special interest in the context of arousal and the physiological assessment of pilot workload for it is this part of the nervous system which controls the heart, secreting glands, and the involuntary muscles. As the name implies control is independent of conscious thought, although exceptions do occur. The main activity of the ANS is concerned with the maintenance, or restoration, of the most favourable internal conditions despite varying demands on the body and despite changes in the external environment. This phenomenon, which involves a series of complex regulatory mechanisms, is known as homeostasis.

The ANS can be divided further into the parasympathetic nervous system (PNS) and the sympathetic nervous system (SNS). Activity in parasympathetic nerve fibres causes the release of acetylcholine, a chemical transmitter, at nerve endings; hence the alternative name for the PNS is cholinergic nervous system. Nor-adrenaline is the chemical transmitter released at the ends of sympathetic nerves and the SNS is sometimes referred to as the adrenergic nervous system. Nor-adrenaline is also secreted, together with a closely related hormone called adrenaline, from the medulla or central part of the adrenal gland. When released into the circulation these hormones, known collectively as catecholamines, augment the activity of the sympathetic nervous system. Because responses to SNS activity are similar to those of adrenal activity, the SNS is sometimes known as the sympathico – adrenal system. In general this system is associated with emergency states, for example, widespread activity occurs in conditions of physiological stress, in situations of danger and under strong emotional stimulii; Cannon[10] described this reaction as preparation for "flight or fight".

The relationship between catecholamines and physiological and psychological states has been the objective of many studies. Frankenhauser and her colleagues[11] have shown that high adrenaline excretion is positively related to better performance in tasks involving perceptual conflict, choice-reaction, and under stimulation. Patkai[12] observed that "... adrenaline seems to be associated with a state of general arousal whereas nor-adrenaline may be related to mechanisms concerned with focussing attention upon specific features of a complex stimulii reaction Studies reported by O'Hanlon[13] showed that adrenaline concentrations were lowered when vigilance levels were reduced, whereas nor-adrenaline concentrations were unaffected. Other workers have shown that in general adrenaline seems to be associated with anxiety and nor-adrenaline with aggression.

ANS activity is normally under the control of centres in the brain (medulla, hypothalamus and cerebral cortex), though some complex responses are controlled by mechanisms at spinal cord level. Emotions, such as fear, anger and grief, affect the ANS via pathways which originate in the cerebral cortex.

A further reference to the adrenal gland, which is part of the endocrine system, is relevant to this section on physiology. The outer part or cortex of the adrenal gland, unlike the medulla, has no nerve supply. Control is by the action of hormones released into the circulation by another endocrine organ, the Pituitary gland, which in turn is controlled by a part of the mid-brain known as the hypothalamus. The adrenal cortex, together with the pituitary and the hypothalamus (hypothalamic – pituitary – adrenal system) is concerned with modulation of behaviour[14] and the response to stress[15]. Hormones secreted by the adrenal cortex, (corticosteroids) are found in various body fluids and, together with the estimation of catecholamines, form the basis of the biochemical methods for assessing workload, stress and fatigue. Because of the difficulty in sampling, these techniques are of much more value in assessing long term effects.

3.3.2 Bioelectric Potentials

A brief note about bioelectric potentials may be of interest because of their relevance to such measures as the electrocardiogram, electroencephalogram and electromyogram.

Many living cells, especially those of nerve and muscle tissue, exhibit a resting electrical potential across their semi-permeable enveloping membrane. When the cell is excited a reversible change occurs and the resting potential is transformed into an action potential of approximately 20mV positive. The mechanism of changing from the resting to the active potential is known as depolarization and the reverse as repolarization. Action potentials are propagated through living tissue by excitation of neighbouring cells. Bioelectric potentials are usually measured by using two or more surface electrodes applied to the skin or occasionally by needle electrodes inserted directly into the tissues. After suitable amplification the bioelectric signals may be recorded onto FM tape, displayed on a cathode ray tube (CRT), or traced on paper by some form of chart recorder.

3.4 PHYSIOLOGICAL VARIABLES

3.4.1 Classification

The most convenient classification of those physiological variables of interest in the assessment of pilot workload is according to functional systems:

A Cardiovascular System – Heart rate
Heart rate variability (sinus arrhythmia)
Blood pressure
Peripheral blood flow
Electrical changes in skin

B Respiratory system – Respiratory rate
Ventilation
Oxygen consumption
Carbon dioxide estimation

C Nervous System – Brain activity
Muscle tension
Pupil size
Finger tremor
Voice changes
Blink rate.

Monitoring techniques used for assessing workload in real-flight must be compatible with flight safety and should be non-intrusive. If used routinely, sensors should be capable of easy and rapid application and, of course, must be acceptable to pilots. These restrictions severely limit the number of physiological variables that can be used in practice, although on an experimental basis it may be possible to employ less than ideal methods. Some of these latter techniques may, with modification and development, become acceptable for routine use in aircraft. A much larger number of physiological indices have been used during studies carried out in laboratories and simulators and these are included in this chapter for completeness.

3.4.2 Cardiovascular

Cardiovascular variables of interest are heart rate, blood pressure, peripheral blood flow and changes in electrical properties of skin, all of which are under the control of the autonomic nervous system and centres in the brain. Although the heart has its own pace-maker made up of a collection of cells known as the sino-auricular node, sympathetic and parasympathetic nerves by acting synergistically exert an overall controlling influence. Stimulation of sympathetic nerves causes acceleration, whereas parasympathetic action is one of inhibition causing a slowing of the heart.

The cardiovascular system is responsible for providing an adequate supply of blood to various tissues and organs of the body. Blood flow can be varied according to local need by changing the diameter of blood vessels supplying the area and if necessary by increasing the output of the heart. This latter action is accomplished either by increasing the stroke volume of the "pump", or by increasing the rate of "pumping", or by both, as happens in response to strenuous exercise. Levels of blood pressure are influenced by similar factors and vary according to the output of the heart and the resistance in the peripheral circulation.

The cardiovascular control centres in the mid-brain are sensitive to stimulation from higher centres in the cerebral cortex; because of these connections, emotional stress such as fear, anger, and excitement can affect the heart rate.

Levels of arousal are indicated by the amount of autonomic system activity which in turn causes cardiovascular changes; these can be measured by suitable monitoring techniques[16].

Heart Rate

Heart rate is one of the easiest of all physiological variables to record and it is a simple matter to obtain precise values because of the discrete signals available in the form of heart beats. It is not surprising that this measure is used so much by research workers interested in behavioural responses. Heart rate can be obtained directly by measuring the heart action or indirectly by counting the arterial pulse which in the healthy person is synchronous with heart rate. Direct measurement techiques consist mainly of sensing the electrical potentials associated with each heart beat – the electrocardiogram, or by detecting the heart sounds with a microphone – the phonocardiogram. Indirect techniques usually involve some means of detecting changes in peripheral blood flow. Normal resting heart rates vary according to fitness and age but are generally in the range 65 to 75 beats per minute (bpm).

Of all the physiological indices available for use in studies of human behaviour heart rate is by far the most popular and it has been measured in a wide variety of situations[17,18,19,20]. Heart rate, recorded in one way or another, has been measured in flight more often than any other physiological variable but few studies have been specifically concerned with assessing pilot workload. Nevertheless, several authors have reported good agreement between heart rate values and degrees of flight task difficulty and there is sufficient evidence to support the practical use of this index[21,22,23].

A more comprehensive description and discussion of measuring heart rate follows in section 3.5.1.

Heart Rate Variability

The heart rate of the normal healthy person at rest varies over periods of seconds by up to 15, or more, beats per minute. This physiological variation in heart rate, known as sinus arrhythmia, is more pronounced in younger persons. It is caused by complex feedback mechanisms associated largely with respiration but also with control of blood pressure and with regulation of skin temperature[24]. Heart rate variability decreases when the mental load is increased and this effect has been studied extensively by Kalsbeek and Ettema[25,26,27] who, using both auditory and visual binary choice tasks, found good correlation between decreased heart rate variability and increased mental load. Similar experiments have been performed by other workers and though most have confirmed the relationship, some doubt has been expressed as to its exact nature.

Several authors have discussed the possible use of sinus arrhythmia as a means of estimating mental workload during complex tasks[28,29]. Mobbs and his colleagues[30] evaluated sinus arrhythmia as a measure of mental workload for possible use in industry; they were unable to demonstrate a consistent relationship and therefore concluded that "... it is quite implausible at this stage to attempt to use heart rate irregularity in the industrial setting". Although sinus arrhythmia cannot be recommended as a measure of pilot workload at present, it does appear to be a sensitive indicator of changes in mental activity and may be used in this role. It will be discussed further in section 3.5.2.

Blood Pressure

Arterial blood pressure is another physiological variable which is frequently used in clinical medicine as a valuable indicator of cardiovascular fitness.

The pumping action of the heart maintains the circulation of blood at a pressure which varies in a pulsatile manner between systolic and diastolic levels. The systolic pressure is that existing in the artery during the heart's contraction, while the diastolic pressure is that during the phase of relaxation. These pressures, being modified by the elastic walls of the arteries, tend to vary inversely with the distance from the heart. The normal systolic pressure is about 120 to 150 mm Hg and the diastolic pressure about 70 to 90 mm Hg, both tending to increase with age.

Laboratory studies have demonstrated the value of blood pressure measurements to indicate levels of arousal and mental activity and several investigators have measured blood pressure on the ground before and after flight as an indicator of stress and fatigue[31,32,33]. Melton et al[34] measured blood pressure, along with other physiological indices, to assess Air Traffic Controllers responses to workload and stress.

In-flight blood pressure measurement has been carried out during a number of studies[35,36]. For example, Roman[37] noted that increases in blood pressure were frequently seen when heart rates and respiratory rates remained low and that blood pressure correlated reasonably well with pilots estimates of task difficulty.

Blood pressure appears to be a promising index of mental workload and stress but present techniques of measurement are not really suitable for routine use. The value of this variable will no doubt improve with further development and also with the collection of more information from flight trials; it will, therefore, be considered further in 3.5.4.

Peripheral Blood Flow

Peripheral blood flow is regulated by a control centre in the brain, the vasomotor centre, via the autonomic nervous system. Increased flow occurs when blood vessels dilate (vaso-dilation) and, conversely, vaso-constriction causes a reduction in blood flow. These changes cause variations in blood pressure as well as affecting the supply of blood to the muscles of the arms and legs. Blood flow to the skin is under the control of a temperature-regulating centre in the brain concerned with maintaining a constant deep body temperature (homeostasis).

For many years, variations in peripheral blood flow have been associated with changes in levels of arousal, with mental activity, and with different emotional states. Blood flow has been measured in several studies of such phenomena and also to assess the effects of drugs[38,39].

Blood flow can be measured in any part of a limb but it is usual in behavioural research to determine vascular changes in a finger or lobe of an ear. The volume of a limb or organ changes according to the amount of blood flowing into it and these changes can be measured by means of an instrument called a plethysmograph. There are two main types, the pulse volume or pneumatic type and the photoelectric plethysmograph. Whereas the former type can be calibrated to produce absolute and relative changes, the photoelectric device demonstrates only the direction of the change.

Different types of plethysmographs have been used to measure blood flow changes in active subjects but, because peripheral blood flow is influenced by temperature, carefully controlled and monitored ambient conditions are necessary during studies of arousal and mental activity; in fact, skin temperature alone can be used for the same purpose. Techniques involving measurement of blood flow and skin temperature tend to be restricted to laboratory use. White (personal communication) of McDonnell-Douglas at Long Beach, has found in his studies that peripheral blood flow is more valuable than heart rate or respiratory rate in indicating levels of mental workload.

Photoplethysmography is probably used more often to measure heart rate than to record changes in blood flow *per se* and will therefore be discussed in more detail later.

Electrical Properties of Skin

Variations in electrical properties of skin are known to occur in response to changes in emotional states and arousal levels and have been measured in behavioural research for many years[16].

In response to a stimulus skin resistance, as measured by passing a small direct current between two electrodes, shows a characteristic decrease known as the Galvanic Skin Response (GSR); the pre-stimulus level being called the Basal Skin Resistance (BSR). These responses are associated in some way with an increase in the activity of sweat glands, although overt sweating is not necessary to produce an effect. Certain areas of the body, notably the palms of the hands and soles of the feet, are concerned more with emotional sweating than with temperature regulation and GSR is usually measured with electrodes applied to one of these areas, for example, the palmar aspect of a finger and the ventral aspect of the wrist.

A related property is that of skin potential which can be measured instead of resistance and by passing an alternating current instead of a direct current between the electrodes, it is possible to measure both potential and resistance at the same time. Changes in skin resistance can also be expressed as changes in skin conductance though it is generally accepted that the former measure is quantitatively more accurate. Amplified signals from skin electrodes can be demonstrated quite easily on a suitable chart recorder and such recordings form an important part of the so-called "lie-detector" test.

GSR has been used to detect changes in arousal, to measure stress, as an indicator of mental activity and to assess car drivers in light and heavy traffic[40]. In 1959 Choen and Silverman[41] suggested skin resistance and electroencephalograms as possible in-flight measurements of mental workload. Skin resistance was included in a battery of physiological indices used to detect changes during a compensatory tracking task[42] and to measure pilot stress during landing[43].

Although variations in skin resistance provide a sensitive indicator of changes in levels of nervous activity results are very susceptible to misinterpretation. In addition the choice of units used to measure GSR is controversial, for example, some give a higher level of GSR and a lower BSR and other units give the opposite. For some years it had been accepted that palmar sited electrodes eliminated the need for ambient temperature monitoring or control but there is now evidence to show that palmar sweat glands also contribute to temperature regulation. In this case, for accurate interpretation of resistance changes, monitoring of skin temperature is essential.

3.4.3 Respiration

Respiration is the physiological process primarily concerned with the interchange of oxygen and carbon dioxide between the body tissues and atmosphere. Cellular activity utilises oxygen obtained from air during inspiration and produces carbon dioxide which is removed during expiration. The quantity of oxygen required by the body is determined by the level of activity or metabolism in various tissues; increased demands being met by increasing the rate and depth of respiration. Control is complex and modulated by neural and chemical factors which are mediated through the autonomic nervous system from the respiratory centre in the hind-brain. Connections with the cerebral cortex make it possible to exercise some degree of voluntary control. The healthy person at rest has a respiratory rate of about 12 breaths per minute. Physical activity causes an increase in rate and depth but emotional influences and increased arousal levels normally cause an increase in rate with a decrease in depth. During periods of stress and intense mental effort a phenomenon known as hyperventilation or over-breathing sometimes occurs.

There are few respiratory indices of interest from the point of view of assessing pilot workload; they include measuring respiratory rate, airflow and volume, and estimating oxygen and carbon dioxide.

Respiratory Airflow and Volume

Airflow is measured by a pneumotachograph of which there are several types; for example, one device measures the pressure across an orifice, and another type measures the change in capacitance. The volume of inspired air may be estimated by simply integrating the results of airflow measurement,

Analysis of Respiratory Gases

Ideally, a system for monitoring respiration in a demanding situation would include the measurement of carbon dioxide in expired air for evidence of hyperventilation[44] and of oxygen to estimate changes in metabolism[45]. Estimation of carbon dioxide and oxygen using some form of gas analyser is normally a laboratory procedure but small and semi-portable instruments have been modified for use in flight simulators and in aircraft[42,46].

Respiratory Rate (RR)

Measurement of breathing rate is probably the most useful of the respiratory variables, it is certainly the easiest to record and has been used extensively as an indicator of emotional states, stress, arousal and mental load[47]. Respiratory rate can be measured in various ways but the commonly used methods for continuous monitoring employ impedence, strain gauge, or thermistor techniques.

Respiratory rate, end tidal carbon dioxide, airflow and ventilation were measured by Benson and his colleagues[42] as part of a phychophysiological study of compensatory tracking. With the exception of carbon dioxide estimation. the same variables were monitored in flight by Corkindale and his co-workers[43] to assess pilot stress during landing. Other authors have reported studies in which a number of respiratory indices have been monitored in flight or in simulated flight[35,37].

Breathing rate has been recorded far more often than have the other respiratory indices and there is evidence to suggest that this variable can be a convenient and useful indicator of mental load and stress. Unfortunately, respiratory patterns are interrupted and modified by speech, thereby reducing the value of rate in flight testing and also in some operational situations. Respiratory rate will be considered further in 3.5.3.

3.4.4 Nervous System

The nervous system is the control and communication system of the body; it is normally divided into the central nervous system (CNS) comprised of the brain, brain stem and spinal cord, and into the peripheral nervous system made up of nerve fibres entering and leaving the CNS. The peripheral nervous system may be divided functionally into the voluntary nervous system and the autonomic nervous system (ANS) which was discussed earlier in section 3.3. The voluntary nervous system is made up of sensory nerves which conduct information to the CNS in the form of coded impulses and motor nerves which transmit impulses from the CNS to various voluntary muscles. Activity in any part of the nervous system is accompanied by electrical changes in nerve cells and by the release of chemical neuro-transmitter substances at nerve endings.

The brain is made up of three developmentally separate parts, the fore-brain, mid-brain and hind-brain. In man, two large structures dominate the rest of the brain; they are the cerebral hemispheres of the fore-brain, and the cerebellum which is part of the hind-brain and which is largely concerned with the subconscious aspects of voluntary movement. All but the superficial grey matter of the cerebral hemispheres, known as the cerebral cortex, is concerned with subconscious or involuntary functions. The cortical areas are connected to each other and to the rest of the brain by various nerve pathways or tracts.

Techniques involving direct measurement of the nervous system are few and the only one of interest in the context of workload is that of electroencephalography and the related phenomenon of evoked potentials. The effects of nervous system activity on other parts of the body are exhibited in most physiological measurements and it is therefore convenient to consider some of them in this section.

The Electroencephalogram (EEG)

EEG potentials represent the combined effect of nerve cell potentials over a large area of the cerebral cortex. They are detected by two or more surface electrodes placed in contact with the scalp or from needle electrodes inserted into the skin. Clinical EEGs are usually derived from surface electrodes placed in a standard pattern thereby allowing valid comparisons to be made of tracings obtained from different subjects.

Although the normal EEG consists of many different frequencies, one usually predominates. An arbitrary classification based on various frequency ranges and using Greek letters is used to describe the rhythms. For example, a commonly described range is the alpha rhythm of about 9-13 Hz; this rhythm is most noticeably affected by visual inputs, the alpha rhythm predominating when the eyes are closed and becoming much less significant when the eyes are opened. Under anaesthesia the alpha rhythm is replaced by the beta rhythm of 14-30 HZ.

In behavioural research the EEG has been widely used in trials into the effects of new drugs, vigilance, mental activity, fatigue and sleep. Experiments have shown that mental activity affects the frequency of the EEG but its value is reduced by large inter- and intra-individual variations. High arousal states are characterised by desynchronisation of EEG signals.

EEGs have been recorded in flight by several investigators[48,49,50] and as a large proportion of the total pilot workload is due to mental effort, this variable would appear, at first sight, to be highly relevant, but unfortunately results are frequently ambiguous and difficult to interpret. Nevertheless, because of the possible promise for the future, this technique will be referred to again in Section 3.5.5.

Evoked Potentials

External stimulii such as intermittent noises or flashing lights evoke a measurable response in the electroencephalogram. Many of these evoked potentials are of low amplitude but, unlike the conventional EEG waves which unfortunately tend to mask them, they are repeatable for similar stimulii and their occurrence can be predicted. By use of suitable summation techniques the signal to noise ratio can be increased so that the evoked response can be readily identified and measured. Digital computers are now commonly used to produce evoked response results direct from the EEG in a readable form.

Evoked potentials have been measured in studies of vigilance and attention or expectation[51,52] and Defayelle and his co-workers[53] described an experiment in which evoked potentials to flashing lights were used to quantify mental load.

Spyker and his colleagues[54] investigated the possible use of the EEG with evoked potentials as a measure of pilot workload but decided that it was unsuitable for this purpose. Groll-Knapp[55], referring to evoked potentials and the nervous system, remarked that "One advantage of the brain potential studies over other phsyiological methods seems to be that we are dealing with a central rather than a peripheral component." However, she concluded: "Brain potentials studies in relation to psychological phenomena and problems are interesting and provocative. But the studies need rather complicated technical equipment, a thorough neurophysiological methodology and a precise and systematic experimental design".

Critical Fusion Frequency

A flashing light will be perceived by the eye as a steady light if the frequency of the flash is increased beyond a certain level. This level is known as the critical fusion frequency (DFF) and varies according to the state of the nervous system.

CFF has been used by research workers as an indicator of fatigue and as a psychophysiological measure during studies of arousal and mental load[56]. Its value in pilot workload studies is obviously limited to use during experiments in laboratories and in flight simulators.

Muscle Tension and Electromyography

The degree of resting tension or tone in different skeletal muscles or groups of muscles depends largely on the attitude of the body and the maintenance of position or posture. Movement and the use of force are accompanied by increased tension in the active muscle groups and a decrease in the passive groups. These changes in tension are reflected by changes in the electrical activity which accompanies muscle fibre contraction. Measurement of this activity is called electromyography (EMG). The EMG can be recorded by surface electrodes placed on the skin over the muscle or by inserting needle electrodes directly into the muscle itself. Signals contain very high frequencies but by applying integration to the complex waveform meaningful results can be obtained.

Electromyography is an important diagnostic and prognostic tool in clinical neurology but, like the EEG, recordings are difficult to interpret. In behavioural studies, it is common practice to measure the EMG in inactive muscles; in 1921 Golla[4] suggested that the magnitude of irrelevant muscle activity is determined by the effort required of relevant muscles in carrying out a set task. EMGs have been used to indicate levels of anxiety and fatigue, to measure reaction times, and to detect and measure tremor[57]. Schnore[16] showed good correlation between arousal levels and physiological measures which included EMGs from muscles of the neck. Duffy[5], who stated that muscle tension and electrical resistance of the skin are undoubtedly related to each other, also pointed out that muscle tension seems to be more consistent than cardiovascular measures.

Lundervold[58] suggested using EMGs to differentiate between potential fighter pilots and potential bomber pilots. He thought that subjects with shorter reaction times and higher muscle potentials would make better fighter pilots. McDonnell[59] reported a simulator study concerned with assessing aircraft handling qualities in which he found a negative correlation between pilot ratings and muscle tension in the active arm, (measured by EMGs). Integrated EMGs from leg and arm muscles were used by Corkindale et al[43] during an in-flight study of pilot stress. Williams and his colleagues[60] used a strain gauge fitted to an aircraft's control column to measure muscle tension during studies of arousal and stress in trainee pilots. They reported an increase in grip pressure on take-off and landing and during solo flight when compared with dual flight.

Wisner[61] has pointed out "... surface EMG is easily recordable but can only be analysed during an experiment in which conditions remain very similar". This comment underlines one of the difficulties of using electromyography to assess pilot workload in aircraft; and it is also difficult to envisage an irrelevant group of muscles in pilots as most muscles are involved at some time or other during the various activities associated with the task of flying an aeroplane.

Physiological Finger Tremor

Most normal subjects exhibit a fine tremor of the outstretched fingers. During emotional states such as excitement, anger and fear the tremor becomes much more obvious. A noticeable tremor, which is made worse by actions requiring fine muscular control such as in writing or in raising a full cup of coffee to the mouth, follows physical exercise or a task demanding a high level of arousal. This phenomenon can be reproduced in experimental subjects by injecting adrenaline into a vein.

Nicholson and his colleagues[62,63] recorded the tremor of an airline pilot by using a strain gauge accelerometer attached to a finger of the outstretched hand. Frequency and acceleration of the tremor were recorded before leaving the ramp (as a base-line) and as soon as possible after landing, as a measure of workload experienced during the preceding let down, approach and landing. In addition to finger tremor, heart rate, or more precisely R-R interval, was recorded during the final stages of the flight. From the results of many flights, these authors concluded that the tremor was indicative of untoward events complicating the approach, whereas heart rate was considered to be more indicative of workload levels during the approach and landing.

Speech Analysis

The characteristics of a person's voice vary when he is subjected to emotional stress. Williams and Stevens[64] carried out an exploratory study on the speech of pilots during flight; they considered that the emotional state was reflected in a number of different acoustic characteristics of the speech signal. Simonov and his colleagues[65,66] analysed the voice frequencies of pilots, cosmonauts and actors, then compared the results obtained during different emotional states and attempted to identify stress and fatigue. Heart rate, which was also recorded, appeared to show some relationship to the voice frequency patterns.

This technique holds some promise for retrospective assessment of stress and workload in accident investigations using speech from cockpit voice recorders, but it seems unlikely to be of use in the day-to-day estimation of workload.

Pupillography

The size of the pupil of the eye varies according to the amount of light shining onto the retina, contracting in bright light and dilating in poor light. It also contracts when looking at near objects as part of the mechanism of accommodation. As well as responding to visual influences the pupil's diameter varies with stress, arousal and mental load[67]. Control of the iris is through the autonomic nervous system and is entirely involuntary. Contraction due to visual inputs is associated with an increase in parasympathetic activity, whereas dilatation is associated with a reduction in parasympathetic activity. The pupil of a frightened person dilates due to sympathetic stimulation, even in the presence of bright light.

Methods of measuring pupil size are mostly based either on photoelectric or on photographic techniques. The latter tend to be somewhat tedious and though the development of infra-red photography has made it possible to record pupil size in very poor light, these techniques are not commonly used. Photoelectric cells, which measure the amount of light reflected from the iris, can provide on-line results and by using image intensifiers pupil changes can be recorded in complete darkness.

Measurement of pupillary diameter is of clinical interest in the diagnosis of diseases of the nervous system but in recent years it has also played an important part in psycho-physiological studies of vision and in experiments into behaviour[68]. Peavler[69] measured pupil sizes of subjects during a short term memory task and observed a significant correlation between individual differences in diameter and recall performance.

Westbrook and his colleagues[1] studied the relationship of pupil size to the difficulty of a manual tracking task and by using a secondary task as a 'conventional' measure of workload, they compared the results with pupil diameter. They tentatively concluded that pupil dilation increased when the tracking task was made more difficult and that the amount of dilation was correlated with workload, as measured by the secondary task, and with the difficulty of the primary task.

There is some evidence that pupillography may be a practical measure of workload in carefully controlled and monitored conditions but it cannot be seriously considered for use in aviation.

Blink Rate

Blinking is a normal everyday action of the eyelids which can be easily seen in most people. It is sometimes a reflex action, for example when something touches the eyeball, or it may be entirely voluntary; it is presumably a mechanism for keeping the cornea moist and clean. Blinking continues all the time on an irregular but frequent basis and cannot be suppressed for long by voluntary effort. The frequency of blinking varies considerably but the intervals are usually in the region of from 2 to 15 sec with each blink lasting for 0.2 to 0.4 sec.

The most convenient method for recording blink rate is by electrooculography but photographic and photoelectric methods, used to observe the pupil or to monitor eye movements, also record blink rate.

It has been shown that blink rate varies with the difficulty of a task, irrespective of whether it is a visual task or not; it appears to be related to muscle tension[47,57].

Poulton and Gregory[70] reported that blink rate decreased when a visual tracking task was made more difficult. Holland and Tarlow[71] recorded blink rate during mental arithmetic and during memory tasks of varying levels of complexity and determined that blink rate was low when mental load was high. A positive correlation between blink rate responses to verbal stress, levels of anxiety and muscle tension was reported by Doehring[72]. Blink rate is of doubtful value in assessing workload at present but further experience of the technique, together with experiments in flight, may prove to be worthwhile, especially if it is presented as a bonus when recording eye movements during studies of pilot activity.

3.4.5 Biochemical Methods

Well established techniques for estimating levels of long term pilot workload and stress and for investigating the physiological affects of fatigue involve measuring levels of various biochemical substances present in such body fluids as blood and urine.

A number of glands in the body, called endocrine glands, secrete their chemical products or hormones directly into the blood where they circulate for the purpose of modulating activity in distant organs and tissues. This method of communication forms an important alternative to that of the nervous system. The endocrine glands mostly concerned with stress and workload are the two adrenal glands, though other glands are variously implicated in some way or other. Each adrenal gland is made up of two functionally separate parts, the central medulla and the outer cortex. Hormones secreted by the medulla, adrenaline and nor-adrenaline, known collectively as catecholamines, tend to augment the action of the sympathetic nervous system (SNS) (see Section 3.3). The adrenal medulla, which has connections with the SNS, is stimulated by a wide variety of stress factors such as fear, anger and hypoxia. Unlike the medulla the adrenal cortex is isolated from the nervous system and relies entirely on circulating hormones. Its relation to the autonomic nervous system is an indirect one, via another endocrine gland called the pituitary and via the hypothalamus, an important control centre for autonomic activity, situated in the forebrain; this relationship is known as the hypothalamic – pituitary adreno-cortical system. The adrenal cortex has several functions but an important one is concerned with adaptation to various forms of stress. Its hormones, known as corticosteroids, include a group called 17 hydroxycorticosteroids (17-OHCS) which are known to increase when the body is adapting to increased load or stress. It is generally accepted that corticosteroids reflect long term stress, whereas catecholamines are indicators of short term stress.

Catecholamines, 17OHCS, and a host of other chemical substances associated with endocrine and metabolic functions have been used as indices of stress, workload and fatigue. For example in 1958 Marchbanks[73] estimated urinary 17OHCS levels as a means of assessing the effects of flight stress on the crew of a jet bomber during a 22½ hour mission. These hormones together with other biochemical indices have been measured during a series of investigations into the effects of long duration flights of various kinds by Hale and his colleagues[74,75,76]. Urinary catecholamines estimations were included in a battery of indices used in studies on pilots engaged in storm penetration flights[77] and flying aerial fire fighting missions[78]. Catecholamines and 17OHCS were both measured to estimate levels of stress in fighter pilots[79,80] and in student pilots[81]. Melton and his colleagues[82] devised a universal stress index calculated from the urinary content of cortico-steroids, adrenaline and nor-adrenaline, which can be readily used in studies of workload and stress in air traffic controllers and flight personnel.

For obvious reasons it is difficult to collect blood or urine at frequent intervals during flight and in most studies using biochemical indices specimens have been collected before and after flight. 17OHCS are present and can be easily estimated in saliva secreted by the parotid glands; these are situated above the angle of the jaw and empty into the mouth via ducts sited near the back teeth. In order to overcome the problems associated with obtaining blood and urine samples, Warren and his co-workers[83] developed a technique for collecting parotid fluid at frequent intervals during flight in high performance aircraft. One study involved collecting parotid fluid samples during different phases of flight in a NF-100 supersonic fighter[84]. Although developed specifically for investigating flight stress, this particular technique may also be of value for assessing short term workload.

3.5 PRACTICAL MEASURES

Of the many physiological indices and techniques described in the previous section, only a small number can be considered to be of practical value for assessing pilot workload during real flight and these will now be discussed further. Heart rate, heart rate variability, and respiratory rate are suitable for routine use; the electroencephalogram and blood pressure are variables which may be employed on an experimental basis.

3.5.1 **Heart Rate** Electrocardiography

The electrocardiogram (ECG) is a graphic representation of changes in the heart muscle potential associated with contraction. Biopotentials from the thick walled ventricle usually produce the largest amplitude changes in the ECG waveform, known as the 'R' wave. In clinical electrocardiography the value of the measure depends to a large extent on electrode siting as the shape of the waveform varies with position; internationally agreed standard electrode positions are in general use to facilitate comparison of records.

Research ECGs have been recorded from subjects engaged in such activities as motor racing[85], downhill ski racing[20], driving an express train[86], and while parachuting[18,87]. In 1940 White[88], using a modified clinical instrument, monitored volunteers while they were flying at various altitudes up to 20,000 ft to determine the effects of hypoxia on their ECGs. The development of suitable equipment for recording airborne ECGs resulted in many studies on pilots to assess the effects of various flight stresses. In 1961 Rowan[89] presented results from test pilots flying experimental aircraft. Roman and Lamb[90] recorded ECGs from pilots flying F-100 supersonic fighters, Holden et al[35] monitored subjects flying as co-pilots in T-33 and F-104 aircraft, and Helvey and his colleagues[91] recorded the ECG of one pilot during a flight test programme with a F-105.

Miniaturised analogue tape recorders, small enough to be carried in the pocket of a flying overall, have made it easy to obtain ECGs during flight in most aircraft[92,93]. Transmission of signals by radiotelemetry has reduced the need to carry recorders, ECGs have been routinely telemetered from astronauts in space[94] and this technique has been successfully used to monitor aircraft pilots[95]. Balke et al[31] telemetered ECG signals over distances of up to 75 miles from pilots flying forest fire fighting missions during studies of stress and fatigue.

Despite major advances in ECG recording equipment and techniques, problems still occur far too often, due in most cases to poor electrode application. Richardson et al[96] in reviewing electrode techniques for long term monitoring of astronauts pointed out: "The weakest link in any long term monitoring system is the electrodes". Electrodes must be capable of maintaining good electrical contact in the presence of sweating and movement, but it is important that neither electrode materials nor conductive jelly cause skin sensitivity leading to inflammation with consequent loss of goodwill from the pilot. Careful positioning of the electrodes may be necessary to produce a large 'R' wave or to reduce the number of artefacts caused by movement in underlying muscle and by such equipment as restraint harness. A multiple chest lead ECG or a trial recording using suction electrodes may be useful in determining optimum electrode positions.

Analysis of electrocardiograms both for evidence of changes in waveform and for determining heart rate can be carried out in various ways from straightforward visual inspection and simple counting to sophisticated computer techniques[92,97].

Good quality ECG waveforms are essential for clinical purposes and for research in cardiac physiology but heart rate can be determined from relatively poor ECGs providing that an unambiguous R wave is present. There is some evidence of waveform changes occurring due to stress[98,99] and so there may be advantages in occasionally recording good ECGs, but in most studies for assessing workload it is convenient to record only heart rate. This can be done easily by using the R wave to trigger some form of counter or to produce an audio signal for recording directly onto magnetic tape. Howitt and his colleagues[100] used this latter technique to monitor heart rates of airline pilots flying scheduled services. By using a pulse of 300 Hz and suitable replay filters, they were able to record speech and heart rate on the same track of the tape. The 'R' wave is also used by a device called the Socially Acceptable Monitoring Instrument (SAMI) which employs a sensitive and reversible electrochemical integrator as a data store[101]. A special replay machine provides a numerical read out of the stored charge as a single total of heart beats over a given time period. A three channel version permits three separate totals to be recorded. Bateman and his co-workers[102] used a SAMI to measure heart rate levels of airline training captains during flight and other activities. By using the 'R' wave of the ECG to trigger a cardiotachometer a direct read out of heart rate may be obtained[103].

Heart Sounds and Phonocardiography

Heart sounds can easily be heard through a stethoscope held against the chest and for many years physicians have used this technique to diagnose disease of the heart. Amplification of the sounds using an electronic stethoscope has made it possible to detect, record and analyse them in detail. A record of heart sounds can be made by placing a microphone against the chest over the heart and connecting it to a suitable amplifier. This instrument, which is used clinically to detect abnormal sounds, is called a phonocardiograph. Phonocardiography is not a practical technique for monitoring heart rate in the noisy environment of the aircraft cockpit, although it has been used to record heart sounds of astronauts in space[104].

Peripheral Pulse

A convenient way of recording the peripheral pulse rate, which in the normal person is the same as heart rate, is by using a photoplethysmograph. This consists of a light source and a shielded photoelectric cell, which operates on the principle that the transmissibility of light through tissues varies according to the flow of blood. Photoplethysmographs can be placed against a flat surface such as the forehead, attached to an ear lobe, or applied to a digit. Willis[105] described a photoelectric pulse detector which has been used to monitor heart rate during flight. He considered the system to be superior in many ways to conventional ECG techniques because it eliminated the need for electrodes and their preparation. An earclip photoelectric device was used by Zenz and Mounts[106] to detect the peripheral pulse in an investigation of heart rate changes during work. Bruner and Hohlweck[107] described a photoplethysmograph combined with a thermister which, when attached to a nostril, measures both heart rate and respiratory rate. They planned to use this technique to investigate levels of workload for airline pilots flying on short haul routes.

Whichever technique is used to obtain heart rate the data has to be analysed and presented in a useable format. If individual beats are recorded some form of cardiotachometer can be used to produce either a digital or meter indication of rate, or a plot of rate against time. Readings are normally averaged over a number of beats or over a period of time but by measuring R-R intervals instantaneous or beat-to-beat heart rate can be obtained. Plots of beat-to-beat rate are particularly useful for analysing variability (sinus arrhythmia) and for detecting rapid changes caused by sudden alterations in workload levels. Heart rate values are most often presented as mean rates for specific epochs of time, or for the entire period of a phase or sub-phase of flight such as an approach and landing[22,108]. Thirty second averages seem to be suitable for most studies, physiological variations being smoothed out but with significant changes usually remaining[100,109]. Mean heart rates for longer[110] and for shorter epochs[23,111] have been used in different studies of pilot stress and workload.

Various automatic and semi-automatic techniques have been developed to reduce the time and effort involved in analysing heart rate data[92]. However, it is worth noting that raw data often contains a lot of valuable information which may be lost with these techniques. Visual inspection of individual beat-to-beat plots, for example, can be most informative.

Heart rate has been used in several general studies of mental activity or workload associated with different tasks. Hashimoto[86] recorded heart rates of express train drivers to estimate their workload and Rohmart and Laurig[112] used heart rate, together with other physiological variables, to assess operator effort during different tasks. Wyncherly and Nicklin[113] found a significant difference in heart rates between a group of blind and a group of sighted pedestrians following the same town route. Other authors have reported on similar studies but the value of heart rate as a measure of mental workload in relatively undemanding jobs is still not really clear.

A large number of airborne studies has involved measuring pilot's heart rate and this physiological variable has been recorded in flight more often than any other. But the literature contains few reports of in-flight studies which refer specifically to assessing workload. Roman and his colleagues[23] measured heart rates of two test pilots flying a series of landings with varying degrees of restricted vision, "to seek a correlation between pulse rate and non-physical workload". The ECG was recorded for 75 seconds before and after touchdown and heart rate was calculated for ten 15 second increments with another 15 second period centred on the touchdown. Results did not show any correlation between heart rate and field of view, though it had been assumed that workload "would markedly increase as horizontal visibility was restricted". Nor was there any correlation between field of view and landing error but there was a high level of correlation between heart rate and landing error. It was concluded that on one sortie of landings gusting wind made conditions more difficult and resulted "in both higher workload and larger landing error".

Test pilot's heart rates were monitored as a routine by Roscoe[108,114] during a series of trials to evaluate various types of noise abatement approaches and landings. Results were used to augment pilot's subjective opinions of workload levels associated with the different approach profiles and techniques. Conventional 3° approaches and landings were used as a datum or standard for comparison with experimental types. In order to minimise the effects of such variables as weather several different types of approaches were compared during the same sortie. Heart rate data were obtained from a modified ECG signal recorded on FM tape and then processed to produce beat-to-beat plots, means for consecutive 30 second epochs, and means for each approach and landing. Heart rate levels agreed quite well with the subjective assessment of workload by the pilots.

Hasbrook and his colleagues[115] monitored heart rates of pilots flying simulated instrument approaches in a light aircraft fitted with an experimental flight instrument display. Heart rate was recorded continuously between the outer marker (OM) and middle marker (MM) but calculated for only five discrete 15 second epochs. Glide slope performance and heart rate responses were similar to those for the conventional display and the subjective reactions of the pilot were favourable. It was therefore concluded that the new display, which reduced panel space by 25%, was an acceptable alternative.

Reports of some other studies in which pilot's heart rates have been monitored have referred to the demands and to the difficulty of the flight task. For example, Roman and Lamb[90] found that "pulse rates correlated well with the pilot's estimate of the difficulty connected with handling the aircraft during any one phase of flight". Rowen[89] pointed out that the high heart rates observed in the pilot of the M-2 lifting body were associated with the poor lift/drag characteristics which made particularly heavy demands on pilot skill. High heart rate levels of M-2 and X-15 pilots were referred to by Carpenter[116] who concluded that "heart rate can be used to estimate portions of the flight that the pilots consider to be most demanding". Ruffell-Smith[117] noted that higher heart rates in airline pilots were associated with more difficult approaches and landings. By measuring pilot's heart rates Billings et al[2] were able to demonstrate that helicopters were significantly less demanding to fly when fitted with a hydraulic boost system.

Nicholson and his co-workers[63] reported a high degree of correlation between the subjective assessment of the overall difficulty of landing approaches and the R-R interval around touchdown. Reporting on a flight trial of noise abatement approaches in a BAC VC10, Gordon-Johnson[118] found "in general good agreement between pilot's heart rate levels and their subjective assessment of workload".

Some studies of flight stress involving experienced pilots have provided results which in many ways can be interpreted in terms of workload[119]. For example, Corkindale and his co-workers[43] recorded heart rate, together with five other physiological variables, to assess pilot stress during landing. The flight trial chosen for their experiment was aimed at evaluating a low visibility approach aid in the form of a head-up display (HUD). A balanced programme compared two conditions, clear visibility and fog screen, and two displays, HUD and normal head-down instrument flight. The authors noted that heart rate, which seemed to differentiate between the four conditions, appeared to be the most sensitive physiological measure. In a study of emotional stress of pilots in special flight conditions such as engine failure, Lapa[120] found that the degree of increase in pulse rate was a function of the complexity of the mental problem within a limited period of time. A number of other authors have reported heart rate changes which seems to reflect variations in stress levels associated entirely with the demands of the flight task,[22,31,121,122,123].

Melton and his colleagues[82] developed a series of biochemical stress indices based on several studies of stress and workload in air traffic controllers and in pilots. They found a significant correlation between heart rate and their overall stress index and, in particular, between heart rate and their adrenaline index. Likewise, Debijadji et al[124] showed good agreement between heart rate and sympathoadrenal reaction and piloting experience and type of flight programme. Melton (personal communication) considers heart rate to be the best available measure of short term workload.

Following a preliminary study of flight-deck workload Howitt and his colleagues[100] concluded that: "The continuous record of heart rates would appear to provide a reliable indication of the pilot's state of arousal, or activation and his current workload".

Thus, there is evidence to support the validity of using heart rate to assess levels of workload in flight but it is necessary to be aware of the limitations inherent in using physiological measures.

3.5.2 Heart Rate Variability (Sinus Arrhythmia)

Physiological variations in heart rate occur in the normal person due to the influence on cardiac control mechanisms of respiration, blood pressure, and skin temperature. Simons and Johnson[125] used the term "reflex heart rate changes" to describe variations caused by respiration and other factors and identified them as: "... transient deviations from homeostasis due to internal or external disturbances". Sayers[24,126] selectively analysed frequency patterns in order to isolate the individual physiological components of heart rate variability. Sinus arrhythmia is usually recorded by a non-integrative cardiotachometer and displayed as a beat-to-beat plot of heart rate against time.

It has been clearly shown that heart rate variability or sinus arrhythmia is supressed when a person is subjected to an increased mental load[27,47]. Many research workers have investigated this phenomenon with the object of developing a variability score which can be used to measure mental workload during tasks of varying complexity. There is no one acceptable method of quantifying variability and a number of different methods have been described, from the relatively simple but practical 'irregularity score' of Kalsbeek and Ettema[25], to complicated spectral analyses employing advanced computer techniques[127]. Different scoring techniques produce different values from the same basic heart rate data leading to ambiguous assessments of mental load. For example, one scoring technique may indicate an increase in mental load, whereas another may not indicate any change. One convenient method of scoring heart rate variability is by calculating standard deviation or variance of the interbeat intervals over a given time, or for a given number of beats.

A number of authors have discussed the use of scored heart rate variability as a method of estimating mental work-load and mental stress during tasks involving vigilance, information processing and decision making. Kalsbeek[128] considered the possibility of using it to measure pilot workload; he pointed out though, that the variability may be suppressed by an increase in overall rate as well as by an increase in mental load and that it may be impossible to differentiate between the two effects. Opmeer and Krol[129] monitored inexperienced pilots in a simulator to compare different levels of cockpit workload for different flight tasks. They were able to differentiate, in increasing order of difficulty, between level flight, holding pattern, take-off, and landing approach. Respiratory rate and sinus arrhythmia were more sensitive than heart rate as indicators of mental load. Sinus arrhythmia was used by Strasser and his co-workers[130] in a study of pilot stress and workload and by Kalsbeek[131] during investigations into air traffic control tasks.

Howitt[132] examined a number of R-R interval plots of pilots flying civil transport aircraft and observed that: "... although certainly as mental work increases the R-R variability decreases, the difference between the same man on different days is so great that we have not found it useful to continue with this aspect of heart rate analysis". Roscoe[109] investigated the value of scoring sinus arrhythmia (obtained as a secondary variable during heart rate monitoring of test pilots) as a means of assessing workload during approach and landing trials. He reported that although sinus arrhythmia appeared to be a sensitive indicator of changes in mental load, there was a noticeable lack of consistency in the results. A more optimistic note was given by Winter of NASA Edwards (personal communication) who is of the opinion that sinus arrhythmia will, with the development of a suitable scoring technique, prove to be a valuable measure of pilot workload.

Although sinus arrhythmia appears to be a sensitive indicator of changes in mental activity, it cannot be recommended as a measure of pilot workload at present. However, it is available as a bonus when the results of monitoring heart rate are presented in beat-to-beat form and then simple visual examination may reveal alterations in mental load in the absence of changes in overall heart rate. Occasionally, it is possible to detect a suppression of variability some seconds before there is a significant increase in rate, thereby providing a more accurate timing of an increase in load.

3.5.3 Respiratory Rate

It is a relatively simple matter to record respiratory rates of aircrew in flight. A temperature sensitive transducer which can be placed in an oxygen mask or in the hose, or on the tip of a boom microphone, is a convenient technique for use in aircraft[133]. Air flowing over the sensor, a thermocouple or a thermistor, causes electrical changes which may be recorded as respiratory rate. A device commonly used in laboratory experiments and occasionally in aircraft, is the chest band strain gauge consisting of a length of silicone rubber tubing filled with mercury. Respiratory movements of the chest cause variations in electrical resistance as the tube stretches and contracts. The impedence pneumograph measures the variations in electrical resistance, between two electrodes placed on either side of the chest, caused by respiratory movement. This method, using suitably sited electrodes, has the advantage that it may be combined with recording the ECG.

Respiratory rates of pilots, frequently with an additional variable such as heart rate, have been monitored many times while flying aircraft. In 1945 Kirsch[134] recorded respiratory rates of aircrew during a combat sortie by counting the movements of flow indicators in the aircraft oxygen system. A mercury strain gauge was used initially by Helvey and his co-workers[91] to monitor respiratory rates of pilots flying F-105 aircraft but they subsequently used a thermistor placed in the nasal airflow to eliminate movement artifacts. Roman[37] and Fraser[135] used heated thermocouples placed in oxygen hose connectors to monitor respiratory rates of pilots flying high performance aircraft. Haward[133] used a thermistor attached to a boom microphone during airborne studies of pilot stress, and Bruner and Hohlweck[107] described a thermister combined with a photoplethysmograph for placing in the nostril during studies of pilot workload. Respiratory rate has been telemetered from aircraft in flight during studies of increased and zero G[35] and as part of an investigation into flight stress and fatigue in pilots flying fire suppression sorties[31].

Eichler and his colleagues[136] measured both respiratory rate and heart rate in pilots of gliders and motor driven sports planes; they found respiratory rate to be the more sensitive indicator of intense concentration. Haward[133], who used respiratory rate during a series of studies of behavioural problems in pilots, considered respiration to be the best single measure of stress. Following experiments in a flight simulator, Opmeer and Krol[129], observed that respiratory rate was a better indicator of mental workload in pilots than sinus arrhythmia and heart rate.

Speech modifies respiration but, notwithstanding this effect, rate may be well worth recording during studies of pilot workload. Unfortunately there is not much evidence of how reliable the measure is in practice but there is certainly evidence that increases in workload and stress cause increased respiratory rates.

3.5.4 Blood Pressure

Physicians usually measure blood pressure by inflating a cuff wrapped around the patient's upper arm and connected to some form of manometer. When the cuff pressure is raised to a level above that of the systolic pressure the pulse at the wrist is obliterated, as the pressure is slowly released the pulse returns and at the same time a stethoscope placed over the artery, just distal to the cuff, will detect the sudden onset of characteristic sounds (known as Korotkoff sounds). After a further fall in pressure the Korotkoff sounds change character and then cease altogether; this point indicates the diastolic level. This is an indirect method of measuring blood pressure and instruments based on this technique are called sphygmomanometers. A direct method of measuring blood pressure is by inserting a cannula into an artery and connecting it to a manometer. This technique, which is used in clinical research and in more detailed examinations of the heart and circulation, results in more accurate and continuous readings. By using a small pressure sensing device and a miniature analogue tape recorder, it is possible to measure continuous blood pressure on subjects engaged in routine activities over a period of several hours[137].

Automatic and semi-automatic measurement of blood pressure based on indirect methods have been developed for research and for clinical monitoring. Compressed gas or a small compressor controlled by a programmed pressure sensitive switch, can be used to inflate the cuff. The Korotkoff sounds can be detected by a microphone placed beneath the cuff and over the artery in a similar manner to the physician's stethoscope.

Measurement of blood pressure during activity is made easier if limited to systolic values only; a pressure transducer can then be used to detect pulsation in the artery; this is particularly advantageous in a noisy environment. Systolic pressure may also be recorded by using a small occlusive cuff applied to a finger, pulsation in the tip of the finger being detected by means of a miniature crystal transducer or by a photoelectric sensor.

A number of automatic techniques have been adapted for airborne use although as long ago as 1977 Gemelli[138] recorded systolic and diastolic pressures in flight using an ordinary clinical sphygmomanometer. Kirsch[134] used a similar instrument to measure systolic blood pressure on aircrew during a combat mission. Holden et al[35] measured blood pressure every 30 sec during a study of pilot response to zero and high G in T-33 and F-104 aircraft. They used nitrogen gas to inflate the cuff and a microphone placed over the brachial artery to detect Korotkoff sounds. Roman and his colleagues[36] used engine compressor bleed air to inflate the cuff and, because of cockpit noise and movement artifacts, used photoelectric sensors to detect arterial pulsation. Roman et al[139] carried out an airborne experiment to assess the accuracy of measuring indirect blood pressure in flight by simultaneously recording direct pressure. The subject pilot, with an intra-arterial catheter in situ, flew in the front seat of a F-100 fighter. It was concluded that although the trial highlighted the "inherent limitations" of the acoustic method it was ". . . sufficiently accurate for all applications now contemplated".

Following a series of in-flight recordings Roman[37] observed blood pressure changes to be more sensitive and more closely related to subjective estimates of task difficulty than heart rate and respiratory rate. This variable holds promise for assessing pilot workload but to be of practical value improved measurement techniques, suitable for routine use in aircraft, are necessary.

3.5.5. Electroencephalography

Measurement of brain activity might seem to be a particularly relevant method for assessing mental workload but at present this is not so. However, electroencephalography is included in this section because recent improvements in monitoring techniques have made in-flight measurement quite practical[93,140]. Further development and experience may lead to this important neuro-physiological measure becoming a suitable method for estimating pilot mental workload.

A multi-channel paper recorder is normally used to produce the electroencephalogram (EEG) in readable form but the EEG signals may be recorded and stored on magnetic tape. Clinical EEGs are usually evaluated by visual inspection but for research purposes the EEG signal is frequently digitised for computer analysis.

EEGs have been recorded from pilots in flight by several investigators; Sem-Jacobson and his colleagues[48,141,142] have recorded eight channel EEGs from pilots flying a jet fighter during studies of flight stresses. Results suggested a strong correlation between EEG changes and the ability of the pilot to perform under conditions of increased G, and also with mental stress generated by instrument flight. EEG signals were telemetered from the flight deck to the rear cabin of a transport aircraft during in-flight studies by La Fontaine and Medvedeff[49]. They overcame many difficulties to obtain tracings without too many artifacts and it was possible for the take-offs and landings to be identified from the different rhythms present in the data. In-flight studies of fatigue and physiological response to inter-continental flights have included EEG monitoring[50]. Howitt (personal communication), who has recorded in-flight EEGs during studies of fatigue and its effect on pilot performance during the approach and landing, considers electroencephalography might be suitable for assessing workload during a demanding flight task.

3.6 OPERATIONAL CONSIDERATIONS

3.6.1 Relevance

A survey of the literature on physiological indices of mental activity, stress and workload does not give a very clear picture of their value, especially in relation to the practical assessment of pilot workload in real flight. Some physiological data are particularly susceptible to misinterpretation and for all indices interpretation can sometimes be quite difficult. Physiological indices may be influenced by factors which are quite unrelated to the task such as smoking a cigarette or eating. In general, though, these factors do not intrude to any extent when the task is realistic and reasonably demanding. Not surprisingly, a number of workers who were at one time enthusiastic supporters of physiological measures have turned their attention to other methods of assessing workload. For example, Hoffelt and Gebert[143] having experimented with physiological variables to assess in-flight strain, decided that psychological measures in the form of pilot interviews were more appropriate.

It is worth noting that a large number of studies associated with measuring workload have been carried out in laboratories and in simulators. Chiles[144] queried the relevance of tasks in laboratory studies to real-world situations and some of the difficulties of transferring results from laboratory experiments to real-life were discussed by Chapanis[145], who concluded his remarks thus: "Although the results of laboratory experiments sometimes provide you with ideas and hunches that may be worth trying out in practical situations, you would be rash to generalise naively from laboratory findings to the solution of real-world problems". Howell[146], in a discussion on pilot workload associated with flight in the terminal movement area (TMA), pointed out "*. . . it is vital to treat human factors as only one but nevertheless important* aspect of the operational problems and not to abstract human factors experiments in isolation or for purely academic reasons". There is an important place for laboratory studies and for simulator experiments, especially in developing equipment and methodology but it is most important to collect as much information as possible from flight trials.

3.6.2 Individual Responses

During early studies of emotion and arousal, results of physiological monitoring revealed many anomalies and discrepancies which were eventually found to be caused by the idiosyncratic responses of the experimental subjects. These factors resulted in either, outright criticism of physiological measures in general, or in monitoring several variables. Individual response specificity, a term used by Lacey[147], has been reported by a number of authors following experiments in laboratories.

It has been shown, for example, that a particular stimulus may cause a large increase in heart rate in one person but not in another, whereas muscle tension may fail to change in the former person but show an appreciable change in the latter. Schnore[16] demonstrated that during qualitatively different arousal conditions subjects exhibited idiosyncratic but highly stereotyped patterns of autonomic nervous system activity.

In addition to response specificity some individuals show characteristically larger physiological responses than do others, to the same stimulii. Those that tend to respond in an over – or hyper-active manner are sometimes termed labile reactors, whereas those that respond in an under – or hypo-active way – are called stabile reactors. This reaction is more or less constant for a particular individual though there is evidence that it can be affected by drugs and illness. Individual responses to different flight tasks have been underlined by several authors[33,90,148,149].

Because of the idiosyncratic physiological response to the demands of the flight task it is necessary, in most instances, for each pilot to be used as his own control.

3.6.3 Combined Measures

The desirability of measuring more than one physiological variable has been stressed by a number of authors. In a review of autonomic nervous system activity Darrow[150] criticised the use of pulse rate as the sole measure of emotional states and suggested that if used it should be in conjunction with monitoring blood pressure. Schnore[16] wrote that ". . . whether comparisons are intra- or inter-individual in character, there are significant tactical advantages in employing several physiological measures rather than relying on only one or two". Duffy[5] suggested that: "Groups of measures rather than a single physiological measure appears to afford a more adequate indication of the general state of arousal". The value of measuring more than one physiological parameter has also been underlined by Benson et al[42] and by Spkyer and his associates[54].

These views have been based largely on laboratory studies where it is relatively easy to use a battery of measures and where appreciable changes in levels of physiological activity rarely occur. Jenny and his colleagues[56] investigated operator workload in an information processing task in which heart rate, oral temperature and critical flicker fusion were measured. In summarising their conclusions, they suggested that: "The absence of significant changes in physiological and perceptual – motor/sensory variables in the present studies may well be a result of the lack of true physiological stress in the laboratory situation". Lazerus and his colleagues[151] studied the relationship between autonomic indicators and physiological stress and reported that ". . . as has been suspected for a long time but never adequately demonstrated, different autonomic indicators of stress do indeed rise and fall together as degrees of stress waxes and wanes". They continued: "Such a finding supports also the reasonableness of employing a small number of autonomic or behavioural response variables (or even a single one) in inferring the presence of physiological stress".

A few airborne studies have involved monitoring a number of physiological indices[43,110,111,135], but it is clearly more expedient to use only one. In most instances the responses of a single variable recorded in flight are adequate and heart rate alone has been measured on many occasions[21,23,108,114]. This particular variable has emerged the clear favourite because of the ease with which it can be recorded and analysed.

3.6.4. Stress and Workload

Most of the early in-flight physiological studies were concerned with assessing the effects of physical stress such as increased and zero G[35,36], hypoxia[88], low level high-speed flight[135] and in developing suitable methodology and equipment[37,139]. Taking advantage of the previous experience of monitoring pilots in aircraft, many of the later studies were aimed primarily at measuring the physiological reaction to mental or psychological stress[33,43]. Several physical stresses cause increases in heart rate, blood pressure and respiratory rate but these effects can be identified or excluded during in-flight studies of workload. Similar increases can be caused in laboratory experiments by emotional stresses such as pain, fear or anxiety, and anger. Some authors have attributed the increases in physiological activity in pilots during the take-off and the approach and landing to fear of physical harm, to risk and to danger. The demands of the flight task certainly result in physiological responses which cannot easily be differentiated from those caused by emotional stresses. However, there is much evidence to show that in the experienced pilot who is in current flying practice, risk and the threat of physical harm do not normally affect heart rate[22,78,122]. Responsibility and paced mental activity are, on the other hand, two psychological stresses which are very closely associated with the task of flying an aircraft and can therefore be considered part of workload[21,117,119]. High workload levels generated by demanding flight tasks are stressful to the pilot and are associated with increased levels of nervous system arousal and preparedness which are reflected in increased physiological activity. In other words, in the competent pilot, physiological responses are normally due to workload and not to unrelated emotional stresses.

3.7 PRACTICAL USE

3.7.1 Applicability

There is little direct evidence available to indicate the real value of physiological measures in assessing pilot workload. Carefully controlled laboratory and simulator experiments, designed to evaluate indices of workload and task difficulty, such as those by Spyker and his colleagues[54], Opmeer and Krol[129], and Soliday and his co-workers[152], appear to have only limited value. However, a large amount of indirect evidence obtained from airborne studies tends to support the validity of measuring physiological variables to assess workload in real flight. Heart rate has been measured in flight more often than any other physiological variable and most of the evidence refers to this index (section 3.5.1). Roman[37] reported that blood pressure responded to changes in task difficulty, as estimated by the pilot, better than did heart rate and respiratory rate. Lauschner and Kirchhoff[153] noted that pulse rates and blood pressures of helicopter pilots reacted in the same sense and with similar amplitudes "as well to psychic stress as to physical workload". Changes in inspiratory minute volumes[154] and in respiratory rate[133] have also been shown to relate to flight stress and task difficulty. In 1969 Howitt[132] stated that for assessing immediate workload". . there is now evidence to suggest that before long it will be possible to use physiological measurements to assess the pilot's level of arousal in terms of those which are optimal for the particular flying task".

Physiological indices of arousal, stress and workload do not result in absolute values and should be used only to indicate relative values. Burger[155], in referring to problems associated with heart rate as a measure of workload, suggested it would be more relevant to use it as a comparative measure and other authors have made similar observations[37,132,156]. Hasbrook and his colleagues[115] measured heart rate in flight to compare two types of aircraft instrumentation, and Roscoe[108,114] used the same variable to compare different types of noise abatement approaches.

3.7.2 Reliability

Physiological responses to a particular flight task may sometimes appear to vary from sortie to sortie, careful examination of the experimental conditions usually shows that extraneous influences, such as weather and air traffic, have changed and therefore a direct comparison cannot be made. Occasionally it is practicable to change the experimental variable during flight in which case it is advantageous to compare two or more variables during the same sortie. It should then be possible to design a series of experimental flights so that results can be evaluated in a statistical manner, thereby improving the reliability of the technique.

The degree of consistency for physiological responses to the same flight tasks or workload levels is difficult to assess because it is virtually impossible to reproduce the exact flight conditions. Roman[36] reported that heart rate, respiratory rate and blood pressure responses were highly reproducible in similar in-flight situations in the same individual. Consistent heart rate values for individual helicopter pilots were noted by Hagelston and his colleagues[121] and Roman et al[23] found consistent heart rate increases, when compared with base-line levels, during experimental landings with restricted vision. Roscoe[148] monitored heart rates of several test pilots flying different types of aircraft and demonstrated reasonably consistent levels for particular pilots, aircraft, and tasks. He observed that the response consistency improved as the task became more demanding and the resulting heart rate levels increased. Physiological measures to assess levels of workload when evaluating handling qualities, systems and procedures are more reliable if the flight task is realistically demanding; and airborne measurements tend to be more reliable than those made in simulators.

3.7.3 Sensitivity

Physiological measures in general have been criticised for being either too sensitive or not sensitive enough. It is sometimes assumed that there is a difference in task difficulty and that physiological measures have failed to detect it when in fact no difference exists at all. An ideal physiological index should be sensitive enough to reveal significant differences in workload levels but not so sensitive that unrealistic differences are indicated.

Physiological variables have been used to differentiate between different levels of workload. For example heart rate clearly differentiated between landing approaches flown in varying weather conditions and between different noise abatement approach techniques[114]. Sinus arrhythmia is a more sensitive variable than are heart rate and respiratory rate but when quantified is inconsistent and unreliable. However, it is of value in identifying changes in mental workload which are not sufficient to cause changes in overall heart rate[109].

3.7.4 Acceptability

Not only must pilots willingly accept being monitored during flight but it is a distinct advantage to have their active cooperation. This means that measuring techniques must be non-intrusive and compatible with flight safety. Sensors should be capable of rapid and easy application without causing discomfort and in general, those used for monitoring heart rate and respiratory rate obey these criteria. Occasionally chest electrodes for detecting the ECG have been left in situ for many hours, having been overlooked by the subject. It is a simple matter to attach disposable electrodes to the chest and routine monitoring of test pilots heart rates can be simplified by pilots applying their own electrodes and connecting various leads before flight. Photoplethysmograph pulse rate sensors are even more easily applied to a finger, ear lobe, or nostril and transducers for measuring respiratory rate do not need to be attached to the person. On the other hand, devices for measuring blood pressure in flight are more likely to be intrusive and also depend upon pre-and post-flight calibration. Application of EEG electrodes to the scalp requires some time, though a special helmet to reduce preparation time has been described[140].

3.7.5 Datum or Base-Line Measurements

Physiological indices do not measure absolute levels of arousal, stress and workload but only relative levels and unless it is possible to compare two or more experimental variables, if possible during the same flight, some form of data or standard is necessary. Roman and Lamb[90] noted that each pilot had his own characteristic level of heart rate for certain conditions of flight and suggested that baselines must be established individually; other authors have similarly stressed the need for measuring individual baseline values. Bateman et al[102] recorded self-counted awakening pulse rates and considered these results to approximate closely to true basal rates. So called 'resting levels' have been recorded during the 'relaxed state' before, during and after simulated and real flight[62,110,111]. An in-flight resting level or datum can be measured during a relatively undemanding part of a sortie when the subject pilot is inactive. A number of studies have used the end of the downwind leg of the circuit pattern as a convenient time during sorties of approaches and landings[21,43,115].

However, Roscoe[148] found this form of baseline was easily influenced by irrelevant stimulii and was therefore too inconsistent and unreliable to be of value, especially when compared with the consistent responses generated by the experimental task itself.

An in-flight or flight task mean level of physiological activity has much to recommend it as a baseline, especially if there is a gradual reduction in the response level throughout the sortie, or part of the sortie, due to a lessening of arousal. A convenient standard may sometimes be available, for example, a 3° instrument approach was used as a datum or standard for comparison purposes in a series of flight trials of steep gradient approaches[109,114].

3.7.6 Familiarisation and Fatigue

Recordings of physiological activity from pilots flying a series of similar tasks frequently show a reduction in response followed by a levelling out as the sortie progresses. This effect is due to familiarisation, learning or adaptation and has been described by several authors[21,91,135]. Even when test pilots have considerable experience of a task the first run of a sortie tends to result in a higher response as the pilot evaluates the effects of weather conditions. In trials flown according to statistical design it is better to exclude the first run from the final analysis.

Occasionally there is an overall and gradual decrease in physiological activity throughout the entire flight which seems to be peculiar to some pilots; this idiosyncratic phenomenon, which is unrelated to familiarisation, appears to be due to an initial over-arousal followed by a slow adaptation.

Rarely towards the end of a sortie, especially if long and demanding or if preceded by others, an increase in physiological activity is evident. This is apparently due to the onset of fatigue when extra effort may be necessary in order to maintain the same level of performance[148].

By designing a flight trial so that experimental sorties can be flown in a statistical manner, the effects of familiarisation and fatigue can be minimised.

3.7.7 Results

Sorties to evaluate handling qualities and workload are rarely flown in identical or ideal conditions; weather, competing traffic, and air traffic control vary from day to day. Certainly, the carefully controlled experimental conditions met with in simulators and laboratories are not available. Moreover, because of the high cost of operating aircraft, the number of flights is usually limited. These constraints make it difficult to obtain statistically significant results and it may be necessary to be content with practical significances and trends.

3.7.8 Performance Monitoring

It is well known that for a specific flight task pilot effort or workload and the resulting performance, are closely related. It is therefore essential that performance should be monitored and acceptable limits clearly defined so that changes in physiological activity can be related to workload and not to variations in performance.

3.8 SUMMARY AND CONCLUSIONS

The rationale of recording physiological activity to assess levels of pilot workload depends on two assumptions: (a) that an acceptable concept of workload is the physical and mental effort required to satisfy the demands of the flight task. And (b) that the level of arousal, as measured by physiological indices, is related in some way to the amount of effort.

Of the various physiological indices heart rate has been shown to be generally reliable for realistically demanding flight tasks and it is reasonably easy to record and to analyse. An added advantage of this measure is that when displayed in beat-to-beat form, heart rate variability is available (as a bonus) for use as a sensitive indicator of changes in mental load.

Because of the limitations inherent in using physiological measures to assess pilot workload there are several pitfalls for the unwary. The following points are worthy of note:

1. Each pilot should normally be used as his own control, thereby minimising the effect of the individual nature of his response.
2. As physiological measures are most valuable when used in a comparative manner, some form of datum or standard is necessary.
3. Comparison is made more meaningful if the experimental condition can be compared with the standard during the same flight.
4. When possible the flight task involved in the assessment of workload should be realistically demanding.
5. Performance should be monitored.
6. Physiological measures appear to be more reliable when the pilot is actually handling the aircraft, i.e. when he is in the aircraft control loop.

Physiological measures alone can be used to estimate levels of workload and especially to identify peaks and troughs in the workload patterns. However, they are of more value when used to augment pilot opinion and, therefore, should be used in conjunction with some form of subjective measure, (see Chapter 2).

REFERENCES

1. Westbrook, C.B.; Anderson, R.O.; Pietrzak, P.E. — *Handling Qualities and Pilot Workload.* In: Conference Proceedings No.14, Assessment of Skill and Performance in Flying. AGARD, Paris, 1966.

2. Billings, C.; Bason, R.; Gerke, R.J. — *Physiological Cost of Piloting Rotary Wing Aircraft.* Aerospace Med. 41, pp.256-258, 1970.

3. Little, D.E. / Joy, R.J.J. — *Energy Cost of Piloting Fixed-wing and Rotary-wing Aircraft.* J. App. Physiol. 26, pp.282-285, 1969.

4. Golla, F.L. — *The Objective Study of Neurosis.* Lancet 2, pp.115-122, 215-221, 265-270 and 373-379, 1921.

5. Duffy Elizabeth. — *Activation and Behaviour.* John Wiley & Sons, New York, 1962.

6. Welford, A.T. — *Stress and Achievement.* Austral. J. Psychol. 17, pp.1-11, 1965.

7. Yerkes, R.M. / Dodson, J.D. — *The Relation of Strength of Stimulus to Rapidity of Habit Formation.* J. Comp. Neur and Psychol. 18, pp.459-482, 1908.

8. Davey, C.P. — *Physical Exertion and Mental Performance.* Ergonomics 16, pp.595-599, 1973.

9. Welford, A.T. — *Stress and Performance.* Ergonomics 16, pp.567-580, 1973.

10. Cannon, W.B. — *Bodily Changes in Pain, Hunger, Fear and Rage.* D. Appleton, New York, 1929.

11. Frankenhauser, M. et al. — *Catecholamine Excretion as Related to Cognitive and Emotional Reaction Patterns.* Psychosom Med. 30, pp.109-120, 1968.

12. Patkai Paula. — *Catecholamine Excretion and Performance.* In: Emotional Stress. Levi, L. (Ed) Karger, New York, 1967.

13. O'Hanlon, J.F. — *Vigilance, the Plasma Catecholamines and Related Biochemical and Physiological Variables.* Human Factors Research. Tech. Report 782-2, Goleta, California, 1970.

14. Davidson, J.M. / Levine, S. — *Endocrine Regulation of Behaviour.* Ann. Rev. Physiol. 34, pp.375-408, 1972.

15. Selye, H. — *Stress in Aerospace Medicine.* Aerospace Med. 44, pp.190-193, 1973.

16. Schnore, M.M. — *Individual Patterns of Physiological Activity as a Function of Task Differences and Degree of Arousal.* J. Exp. Psychol. 58, pp.117-128, 1959.

17. Taggart, P. / Gibbons, D. / Somerville, W. — *Some Effects of Motor-car Driving on the Normal and Abnormal Heart.* British Medical Journal, 4, pp.130-134, 1969.

18. Schane, W.P. / Slinde, K.E. — *Continuous ECG Recording During Free Fall Parachuting.* Aerospace Med. 39, pp.597-603, 1968.

19. Barnard, R.J. / Duncan, H.W. — *Heart Rate and ECG Responses of Fire Fighters.* J. Occ. Med. 17, pp.247-250, 1975.

20. Hanson, J.S. / Tabakin, B.S. — *Electrocardiographic Telemetry in Skiers Anticipatory and Recovery Heart Rate During Competition.* New England. J. Med. 27, pp.180-185, 1964.

21. Hasbrook, A.H. / Rasmussen, P.G. — *Pilot Heart Rate During In-flight Simulated Instrument Approaches in a General Aviation Aircraft.* Aerospace Med. 41, pp.1148-1152, 1970.

22. Lewis, C.E. et al. — *Flight Research Program IX: Medical Monitoring of Carrier Pilots in Combat II.* Aerospace Med. 38, pp.581-592, 1967.

23. Roman, J. et al. — *Flight Research Program VI: Heart Rate and Landing Error in Restricted Fields of Vision Landings.* Aerospace Med. 38, pp.128-132, 1967.

24. Sayers, B.McA. — *An Analysis of Heart Rate Variability.* Ergonomics 16, pp.85-97, 1973.

25. Kalsbeek, J.W.H. / Ettema, J.H. — *Scored Regularity of the Heart Pattern and the Measurement of Perceptual or Mental Load.* Ergonomics 6, pp.306-307, 1963.

26. Kalsbeek, J.W.H. / Ettema, J.H. — *Physiological and Psychological Evaluation of Distraction Stress.* Proceedings 2nd International Ergonomics Association Congress. Dortmund, pp.443-447, 1964.

27. Kalsbeek, J.W.H. / Ettema, J.H. — *Sinus Arrhythmia and the Measurement of Mental Load.* Communication at the London Conference of the British Phychological Society, 1965.

28. Kalsbeek, J.W.H. — *Measurement of Mental Workload and Acceptable Load: Possible Application in Industry.* Int. J. Prod. Res. 7, pp.33-45, 1968.

29. Ettema, J.H., Zielhuis, R.L. — *Physiological Parameters of Mental Load.* Ergonomics. 14, pp.137-144, 1971.

30. Mobbs, R.F. et al. — *An Evaluation of the Use of Heart Rate Irregularity as a Measure of Mental Workload in the Steel Industry.* BISRA Report No. OR/HF/25/71, London, 1971.

31. Balke, B. et al. — *Physiological Stress and Fatigue in Aerial Missions for the Control of Forest Fires.* Aerospace Med. 37, pp.221-227, 1966.

32. Melton, C.E. et al. — *Effect of a General Aviation Trainer on the Stress of Flight Training.* Aviat Space Environ Med. 46, pp.1-5, 1975.

33. Blix, A.S. et al. — *Additional Heart Rate – An Indication of Psychological Activity.* Aerospace Med. 45, pp.1219-1222, 1974.

34. Melton, C.E. et al. — *Physiological Responses in Air Traffic Control Personnel: O'Hare Tower.* FAA Office of Aviation Medicine Report No.AM-71-2, 1971.

35. Holden, G.R. et al. — *Physiological Instrumentation System for Monitoring Pilot Response to Stress at Zero and High G.* Aerospace Med. 33, pp.420-427, 1962.

36. Roman, J.A. et al. — *School of Aerospace Medicine. Physiological Studies in High Performance Aircraft.* Aerospace Med. 33, pp.412-419, 1962.

37. Roman, J.A. — *Cardio-respiratory Functioning in Flight.* Aerospace Med. 34, pp.322-337, 1963.

38. Meyer-Delius, J. et al. — *Photoplethysmographic Monitoring of Vigilance Related to Visual Perception.* Paper presented to the 45th Annual Scientific Meeting of Aerospace Medical Association, 1974.

39. Lader, M.H. — *The Effects of Cyclobarbitone on the Pulse Volume, Pulse Rate and Electromyogram.* J. Psych Res 8, pp.395-398, 1965.

40. Preston, Barbara. — *Insurance Classification and Driver's Galvanic Skin Response.* Ergonomics. 12, pp.437-446, 1969.

41. Cohen, S.I., Silverman, A.J. — *Measurement of Pilot Mental Effort.* Report No.148, AGARD, Paris, 1957.

42. Benson, A.J. et al. — *A Psychophysiological Study of Compensatory Tracking on a Digital Display.* Human Factors. 7, pp.457-472, 1965.

43. Corkindale, K.G. et al. — *Physiological Assessment of Pilots Stress During Landing. In: Conference Proceedings No.56.* Measurement of Aircrew Performance. AGARD, Paris, 1969.

44. Murphy, T.M., Young, W.A. — *Hyperventilation in Aircraft Pilots.* Aerospace Med. 39, pp.463-466, 1968.

45. Wald, A., Harrison Lynne, B. — *Oxygen Consumption and Heart Rate: Changes and Relationships in Static Work.* Ergonomics. 18, pp.299-309, 1975.

46. Brigden, W.H., Roman, J. — *Flight Research Program: IV A Small Gas Analyser for Aerospace.* Aerospace Med. 37, pp.1037-1040, 1966.

47. Thackray, R.I. — *Patterns of Physiological Activity Accompanying Performance on a Perceptual-motor task.* FAA Office of Aviation Medicine Report No.FAA-AM 69-8, 1969.

48. Sem-Jacobson, C.W. — *Electroencephalographic Studies of Pilot Stress in Flight.* J. Aviat Med. 30, pp.797-801, 1959.

49. La Fontaine, C.B., Medvedeff, M. — *Radiotelemetric Recordings of the Electroencephalograms of Civil Aviation Pilots During Flight.* Aerospace Med. 30, pp.1060-1065, 1966.

50. Carruthers, M. et al. — *Man in Transit: Biochemical and Physiological Changes During Intercontinental Flights.* Lancet. 1, pp.977-981, 1976.

51. Donchin, E. Cohen, L. — *Averaged Evoked Potentials and Intra-modilaty Selective Attention.* J. Electroenceph Clin. Neurophysiol. 22, pp.537-546, 1967.

52. Groves, M. Eason, R.G. — *Effects of Attention and Activation on the Visual Evoked Cortical Potential and Reaction Time.* Psychophysiology. 5, pp.394-398, 1969.

53. Defayolle, M. et al. — *Averaged Evoked Potentials in Relation to Attitude, Mental Load and Intelligence.* In: Measurement of Man at Work. Singleton, W.T, Fox, J.G. and Whitfield, D. (Eds). Taylor and Francis, London, 1971.

54. Spyker, D.A. et al. — *Development of Techniques for Measuring Pilot Workload.* NASA Contractor Report No. CR 1888. Washington DC, 1971.

55. Groll-Knapp, E. — *Evoked Potential and Behaviour.* In: Measurement of Man at Work. Eds. Singleton, W.T. Fox, J.G. and Whitfield, D. Taylor and Francis, London, 1971.

56. Jenney, L.L. et al. — *Measurement of Operator Workload in an Information Processing Task.* NASA Contractor Report No.CR-2150. Washington DC, 1972.

57. Webb, R.A. Obrist, P.A. — *The Physiological Concomitants of Reaction Time Performance as a Function of Preparatory Interval and Preparatory Interval Series.* Psychophysiology. 6, pp.389-403, 1970.

58. Lundervold, A. — *Electromyography as a Test for Pilot Aspirants.* J. Aviat Med. 21, pp.147-149, 1950.

59. McDonnell, J.D. — *Pilot Rating Techniques for the Estimation and Evaluation of Handling Qualities.* Air Force Flight Dynamics Laboratory Report. TR 68-76, Wright-Patterson AFB, 1968.

60. Williams, A.C. et al. — *Preliminary Experimental Investigation of Tension as a Determinant of Performance in Flight Training.* CAA Division of Research Report No.54. Washington DC, 1946.

61. Wisner, A. — *Electrophysiological Measures for Tasks of Low Energy Expenditure* In Measurement of Man at Work. Singleton, W.T, Fox, J.G. and Whitfield, D (Eds). Taylor and Francis, London, 1971.

62. Nicholson, A.N. et al. — *Activity of the Nervous System During the Let-down, Approach and Landing; A Study of Short Duration High Workload.* Aerospace Med. 41, pp.436-446, 1970.

63. Nicholson, A.N. et al. — *Influence of Workload on the Neurological State of a Pilot During the Approach and Landing.* Aerospace Med. 44, pp.146-152, 1973.

64. Williams, C.E. Stevens, K.N. — *On Determining the Emotional State of Pilots During Flight: An Exploratory Study.* Aerospace Med. 40, pp.1369-1372, 1969.

65. Simonov, P.V. Frolov, M.V. — *Utilisation of Human Voice for Estimation of Man's Emotional Stress and State of Attention.* Aerospace Med. 44, pp.256-258, 1973.

66. Simonov, P.V. et al. — *Use of the Invariant Method of Speech Analysis to Discern the Emotional State of Announcers.* Aviat Space Environ Med. 46, pp.1014-1016, 1975.

67. Hess, E.H. Polt, J.M. — *Pupil Size in Relation to Mental Activity During Simple Problem Solving.* Science. 143, pp.1190-1192, 1964.

68. Poock, G.K. Noel, C.E. — *Effects of Layout of Workplace on Pupillary Diameter* Perceptual and Motor Skills. 40, pp.304-310, 1975.

69. Peavler, W.S. — *Pupil Size, Information Overload, and Performance Difference.* Psychophysiology. 11, pp.559-566, 1974.

70. Poulton, E.C. Gregory, R.L. — *Blinking During Visual Tracking.* Quart. J. Exp. Psychol. 4, pp.57-65, 1952.

71. Holland, M.K. Tarlow, G. — *Blinking and Mental Load.* Psychological Reports. 31, pp.119-127, 1972.

72. Dochring, D.G. — *The Relation Between Manifest Anxiety and Rate of Eye Blinks in a Stress Situation.* Naval. School of Aviation Report No.6. Pensacola, 1957.

73. Marchbanks, V.H. *Effect of Flying Stress on Urinary 17-Hydroxycorticosteroid Levels.* J. Aviat Med. 29, pp.676-682, 1958.

74. Hale, H.B. et al. *Endocrine-metabolic Effect of Unusually Long or Frequent Missions in C-130E or C-135B Aircraft.* Aerospace Med. 39, pp.561-570, 1968.

75. Hale, H.B., Williams, E.W., Buckley, C.J. *Aeromedical Aspects of the First Non-stop Translantic Helicopter Flight. III Endocrine-metabolic Effects.* Aerospace Med. 40, pp.718-723, 1969.

76. Hale, H.B. et al. *Physiologic Cost of Prolonged Double Crew Flights in C-5 Aircraft.* Aerospace Med. 44, pp.999-1008, 1973.

77. McKenzie, J.M., Fiorica, V. *Stress Responses of Pilots to Severe Weather Flying.* Aerospace Med. 38, pp.576-580, 1967.

78. Melton, C.E. et al. *Physiological Studies of Air Tanker Pilots Flying Forest Fires Retardent Missions.* FAA Office of Aviation Medicine Report No.FAA AM-68-26 Washington CD, 1968.

79. Demos, G.T. et al. *Anticipatory Stress and Flight Stress in F-102 Pilots.* Aerospace Med. 40, pp.385-388, 1969.

80. Austin, F. et al. *Aeromedical Monitoring of Naval Aviators During Aircraft Carrier Combat Operations.* Aerospace Med. 38, pp.593-596, 1967.

81. Melton, C.E., Fiorica, V. *Physiological Responses of Low-time Private Pilots to Cross-country Flying.* FAA Office of Aviation Medicine. Report No.FAA AM-69-12 Washington DC, 1969.

82. Melton, C.E. et al. *Studies on Stress in Aviation Personnel: Analysis and Presentation of Data Derived from a Battery of Measurements.* In: Conference Proceedings No.180. The Role of the Clinical Laboratory in Aerospace Medicine. AGARD, Paris, 1975.

83. Warren, B.H. et al. *Determination of Inflight Biochemical Responses Utilising the Parotid Fluid Collection Technique.* Aerospace Med. 37, pp.796-799, 1966.

84. Warren, B.H. et al. *Serial Sampling of Parotid Fluid for 17 – Hydroxycorticosteroid Levels as a Quantitative Estimate of Stress During Flight in High-performance Jet aircraft.* In: Conference Proceedings No.2. Collected papers presented at the 22nd meeting of the ASMP. AGARD, Paris, 1965.

85. Rapp, R.M., Dietlein, L.F., Nuttall, J.C. *Some Physiological Observations and Monitoring Techniques Associated with Auto-racing.* Aerospace Med. 36, 159, 1965.

86. Hashimoto, K. *Estimation of the Driver's Workload in High Speed Electric Car Operation of the New Tokaido Line in Japan.* In: Proceedings of Second International Congress on Ergonomics. Dortmund, 1964.

87. Deroanne, R. et al. *Telemetric Control of Heart Adaptation During Automatic and Free-fall Parachute Jumps.* Aviat Space Environ Med. 46, pp.128-131, 1975.

88. White, M.S. *The Effect of Anoxia in High Altitude Flights on the Electrocardiagram.* J. Av. Med. 11, pp.166-180, 1940.

89. Rowen, B. *Biomedical Monitoring of the X-15 Program Air Force Flight Test Centre Edwards AFB* TN-61-4, 1961.

90. Roman, J.A., Lamb, L.E. *Electrocardiography in Flight* Aerospace Med. 33, pp.527-544, 1962.

91. Helvey, W.M. et al. *A Review of Biomedical Monitoring Activities and Report on Studies Made on F-105 Pilots.* Aerospace Med. 35, pp.23-27, 1964.

92. Carpenter, R., Roman, J. *Recording and Signal-Conditioning Techniques and Equipment Used in a 1000-flight Biomedical Study.* NASA TN D-4487, Washington DC, 1968.

93. McKinnon, J.B. — *A Miniature 4-channel Cassette Recorder for Physiological and Other Variables.* In Biotelemetry II. Neukomm (Ed) Karger Basel pp.67-90, 1974.

94. Luezkowski, S.M. — *Bio-instrumentation* In: Biomedical Results of Appollo. Johnston, R, Dietlein, F.F. and Berry, C.A. (Eds) NASA SP-368, Washington DC, 1975.

95. Barr, N.L., Voas, R.B. — *Telemetering Physiological Responses During Experimental Flights.* American J. Cardiol. 6, pp.55-61, 1960.

96. Richardson, P.C. et al. — *Some New Electrode Techniques for Long-term Physiological Monitoring.* Aerospace Med. 39, pp.745-750, 1968.

97. Cashman, P.M.M., Stott, F.D. — *A Semi-automatic System for the Analysis of 24 Hour ECG Recordings from Ambulant Subjects.* Biomedical Engineering. 8, pp.54-57, 1974.

98. Mitchell, J.H., Shopiro, A.P. — *The Relationship of Adrenalin and T-wave Changes in the Anxiety State.* American Heart J. 42, pp.323-328, 1954.

99. Simonov, P.V. et al. — *Characteristics of the Electrocardiogram Under Physical and Emotional Stress in Man.* Aviat Space Environ Med. 46, pp.141-143, 1975.

100. Howitt, J.S. et al. — *A Preliminary Study of Flight Deck Workloads in Civil Air Transport Aircraft.* Flying Personnel Research Committee. Report No.1240 (London), 1965.

101. Barker, J.A. et al. — *Socially Acceptable Monitoring Instruments (SAMI)* J. Physiol. 188, pp.4-5, 1966.

102. Bateman, S.C. et al. — *Heart Rate of Training Captains Engaged in Different Activities.* Aerospace Med. 41, pp.425-429, 1970.

103. Pope, J.M. et al. — *A Cardiotachometer with Linear Indication of Beat-to-beat Frequency.* Psychophysiology. 4, pp.486-492, 1968.

104. Vallbona, C. et al. — *Effect of Orbital Flight on the Duration of the Cardiac Cycle and of its Phases.* Aerospace Med. 41, pp.529-537, 1970.

105. Willis, D.M. — *A Photoelectric Ear-device for In-flight Recording of Heart Rate.* Biomedizinische Technik. 19, pp.5-9, 1974.

106. Zenz, C., Mounts, F. — *Determination of the Heart Rate During Work.* AMA Arch Indust. Health. 17, pp.280-286, 1958.

107. Brüner, H., Hohlweck, H. — *A New Method for Recording the Heart and Respiratory Rates of Cockpit Crews in Flight.* In: Current Research Work at the Institute for Aerospace Medicine. Bonn-Bad Godesberg Tech. Transl. of DLR-FB 73-15, pp.19-29, 1973.

108. Roscoe, A.H. — *Heart Rate Monitoring of Pilots During Steep Gradient Approaches.* Aviat. Space Environ Med. 46, pp.1410-1415, 1975.

109. Roscoe, A.H. — *Heart Rate Changes in Test Pilots.* Paper presented to 65th Main Scientific Meeting Biological Engineering Societies, London, January 1977. (In press).

110. Melton, C.E., Wicks, Marlene. — *In-flight Physiological Monitoring of Student Pilots.* FAA Office of Aviation Medicine Report No.FAA-AM, 67-15, 1967.

111. Rolfe, J.M. et al. — *Pilot Response in Flight and Simulated Flight.* RAF Institute of Aviation Medicine. Report No.437, Farnborough, 1968.

112. Rohmert, W., Laurig, U. — *Work Measurement Psychological and Physiological Technique for Assessing Operator and Workload.* Inst. J. Prod. Res. 9, pp.157-168, 1971.

113. Wyncherly, R.J., Nicklin, B.H. — *The Heart Rate of Blind and Sighted Pedestrians on a Town Route.* Ergonomics. 13, pp.181-192, 1970.

114. Roscoe, A.H. — *Pilot Workload During Steep Gradient Approaches.* In: Conference Proceedings No.212. Aircraft Operational Experience and its Impact on Safety and Survivability. AGARD, Paris, 1976.

115. Hasbrook, H. Rasmussen, P.G. Willis, D.M. — *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.

116. Carpenter, R. — *Biotechnology Problems Relative to the Space Shuttle Vehicle.* J. of Aircraft. 8, pp.662-666, 1971.

117. Ruffell-Smith, H.P. — *Heart Rate of Pilots Flying Aircraft on Scheduled Airline Routes.* Aerospace Med. 38, pp.1117-1119, 1967.

118. Gordon-Johnson, P. — *The Development of Noise Abatement Approach Procedures and a Pilot's Subjective Assessment of Workload During the Approach and Landing Phase of Flight.* Paper presented to 46th Annual Scientific Meeting of Aerospace Medical Assn. San Francisco, April 1975.

119. Roscoe, A.H. — *Stress, Responsibility and Workload.* Paper presented to 48th Annual Scientific Meeting of Aerospace Medical Assn. Las Vegas, May 1977.

120. Lapa, V.V. — *Study of the State of Emotional Stress of Pilots in Special Flight Situations.* (In Russian). Voenno-Meditsinski Zhumal. pp.65-67. English Abst. In International Aerospace Abstracts 11 44473, Washington DC, 1971.

121. Hagelston, J.O. et al. — *Continuous ECG Monitoring of Helicopter Pilots.* Medica Mundi. 18, pp.87-89, 1973.

122. Roman, J. — *Flight Research Program: II Risk and Responsibility as Factors Affecting Heart Rate in Test Pilots.* Aerospace Med. 36, pp.518-523, 1965.

123. Koetz, H. — *Concerning Pulse Rate Increments of Helicopter Pilots as a Measurement of Their Stress* (In German). Luft-und Raumfahrt-Medizin. 14, pp.65-70, 1968.

124. Debijadgi, R. et al. — *Psychological Reactions of Pilots in Supersonic Aircraft in Flight* (In French). Revue de Médicine Aéronautique et Spatiale. 12, pp.367-371, 1973.

125. Simons, D.G. Johnson, R.L. — *Heart Rate Patterns Observed in Medical Monitoring.* Aerospace Med. 36, pp.504-513, 1965.

126. Sayers, B.McA. — *The Analysis of Cardiac Interbeat Interval Sequences and the Effect of Mental Workload.* Proc. Royal Soc. Med. 64, pp.707-710, 1971.

127. Luczak, H. Laurig, W. — *An Analysis of Heart Rate Variability.* Ergonomics. 16, pp.85-97, 1973.

128. Kalsbeek, J.W.H. — *Objective Measurement of Mental Workload; Possible Applications to the Flight Task.* In Conference Proceedings No.55. Problems of the Cockpit Environment. AGARD, Paris, 1968.

129. Opmeer, C.H.J.M. Krol, J.P. — *Towards an Objective Assessment of Cockpit Workload*: Physiological Variables During Different Flight Phases. Aerospace Med. 44, pp.527-532, 1973.

130. Strasser, H. et al. — *Physiological and Operational State of a Group of Aeroplane Pilots Under the Conditions of Stressing Tracking Tasks.* Aerospace Med. 44, pp.1040-1047, 1973.

131. Kalsbeek, J.W.H. — *Standards of Acceptable Loads in ATC Tasks.* Ergonomics. 14, pp.641-650, 1971.

132. Howitt, J.S. — *Flight-deck Workload Studies in Civil Air Transport Aircraft.* In: Conference Proceedings No.56. Measurement of Aircrew Performance. AGARD, Paris, 1969.

133. Haward, L.R.C. — *Assessment of Stress-tolerance in Commercial Pilots.* Flight Safety. 1, pp.12-17, 1967.

134. Kirsch, R.E. — *A Physiological Study of Aviators During Combat Flying.* J. Aviat Med. 16, pp.376-384, 1945.

135. Fraser, T.M. — *Aspects of the Human Response to High Speed Low Level Flight.* Aerospace Med. 35, pp.365-370, 1964.

136. Eichler, J. et al. — *Telemetry of the Heart Rate and Respiratory Rate in Pilots of Gliding and Motor Driven Sports Planes Under Different Flight Conditions.* (In German), Mschr Unfallheilk. 78, pp.283-289, 1975.

137. Littler, W.A. et al. — *Continuous Recording of Direct Arterial Pressure and Electrocardiogram in Unrestricted Man.* British Medical Journal. 3, pp.76-78, 1972.

138. Gemelli, A. — *On the Application of Psycho-physical Methods of Examining Candidates as Military Pilots.* (In Italian). Paper presented to Congress of the Society for the Advance of Sciences. Vita e Pensiero, April 1917.

139. Roman, J. et al. — *Validity of Flight Blood Pressure Data.* Aerospace Med. 36, pp.436-441, 1965.

140. Westbrook, R.M. Zuccara, J.J. — *EEG Sensing and Transmitting System Contained in a Flight Helmet.* Aerospace Med. 40, pp.392-396, 1969.

141. Sem-Jacobsen, C.W. et al. — *EEG Recording in Simulated Combat Flight in Jet Fighter Planes.* Electroenceph clin. Neurophysiol. 11, p.154, 1959.

142. Sem-Jacobson, C.W. Sem-Jacobson, I.E. — *Selection and Evaluation of Pilots for High Performance Aircraft and Spacecraft by In-flight EEG Study of Stress Tolerance.* Aerospace Med. 30, pp.605-609, 1963.

143. Hoffelt, W. Gerbert, K. — *Some Critical Comments on the Measurement of In-flight Strains.* In Conference Proceedings No.101. Performance and Biodynamic Stress. AGARD, Paris, 1972.

144. Chiles, W.D. — *Complex Performance: the Development of Research Criteria Applicable to the Real World.* In: Measurement of Man at Work. Singleton, W.T, Fox, J.G. and Whitfield, D. (Eds). Taylor and Francis, London, 1971.

145. Chapanis, A. — *The Relevance of Laboratory Studies to Practical Situations.* Ergonomics. 10, pp.557-577, 1967.

146. Howell, G.C. — *Future System Consideration for Operations of Aircraft in the Terminal Movement Area.* Presented at 14th Anglo American Aeronautical Conference. AIAA Paper. 75-957, Los Angeles, 1975.

147. Lacey, J.I. et al. — *Autonomic Response Specificity:* An Experimental Study. Psychosom Med. 15, pp.8-21, 1953.

148. Roscoe, A.H. — *Use of Pilot Heart Rate Measurement in Flight Evaluation.* Aviat Space Environ Med. 47, pp.86-90, 1976.

149. Billings, C.E. et al. — *Stress and Strain in Student Helicopter Pilots.* Aerospace Med. 44, pp.1031-1035, 1973.

150. Darrow, C.W. — *The Galvanic Skin Reflex (Sweating) and Blood Pressure as Preparatory and Facilitative Functions.* Psychol Bull. 33, pp.73-82, 1936.

151. Lazerus, R.S. et al. — *The Relationship Between Autonomic Indications of Psychological Stress: Heart Rate and Skin Resistance.* Psychosom Med. 25, pp.19-30, 1963.

152. Soliday, S.M. Schohan, B. — *Performance and Physiological Responses of Pilots in Simulated Low-altitude High-speed Flight.* Aerospace Med. 36, pp.100-104, 1965.

153. Lauschner, E.A. Kirchhoff, H.W. — *Comparison of Cardiovascular Reactions Originated by Difficult Helicopter Flight Patterns and Physical Workload.* In Conference Proceedings No.2. Collected papers presented at the 22nd meeting of the ASMP AGARD, Paris, 1965.

154. Pettyjohn, F.S. et al. — *Use of Inspiratory Minute Volumes in Evaluation of Rotary and Fixed Wing Pilot Workload.* In: Conference Pre-prints No.216. Methods to Assess Workload. AGARD, Paris, 1977.

155. Burger, G.C.E. — *Heart Rate and the Concept of Circulatory Load.* Ergonomics. 12, pp.857-864, 1969.

156. Benson, A.J. Rolfe, J.M. — *The Use of Psychophysiological Measures in the Assessment of Operator Effort.* In: Conference Proceedings No.14. Assessment of Skills and Performance in Flying. AGARD, Paris, 1966.

BIBLIOGRAPHY

Abramson, D.I. Ferris, E.B.	*Response of Blood Vessels in the Resting Hand and Forearm to Various Stimuli.* American Heart J. 19, pp.541-533, 1940.
Anrep, G.V. et al.	*Respiratory Variations of the Heart Rate 1. The Reflex of the Respiratory Arrhythmia.* Proc. Royal Soc. Med. 119, pp.191-196, 1936.
Balke, B. et al.	*In-flight Hyperventilation During Jet Pilot Training.* J. Aviat Med. 28, pp.241-248, 1957.
Ball, G.R. et al.	*An Automatic Finger-pressure Recorder.* Lancet. 2, pp.1178-1180, 1961.
Blair, D.A. et al.	*Excitation of Cholinergic Vasodilator Nerves to Human Skeletal Muscles During Emotional Stress.* J. Physiol. 148, pp.633-647, 1959.
Boyce, P.R.	*Sinus Arrhythmia as a Measure of Mental Load.* Ergonomics. 17, pp.177-183, 1974.
Burch, G. Ray, T.	*Cardiovascular System as the Effector Organ in Psychosomatic Phenomena.* J. American Med. Assoc. 136, pp.1011-1017, 1948.
Carmichael, E.A. et al.	*A Physiological Study of the Skin Resistance Response in Man.* J. Physiol. 99, pp.329-337, 1941.
Carroll, D.	*Physiological Response to Relevant and Irrelevant Stimulii in a Simple Reaction Time Situation.* Egonomics. 16, pp.587-594, 1973.
Clark, W.A.	*Digital Technique in Neuroelectric Data Processing.* J. Electroenceph Clin. Neurophysiol. 20, pp.75-78, 1961.
Cromwell, L. et al.	*Biomedical Instrumentation and Measurements.* Prentice-Hall, New Jersey, 1973.
Danev, S.G. de Winter, C.R.	*Expectation, Attention and Heart Rate Deceleration – Some Complication in the Use of Heart Rate Variability for Assessing Mental Load.* Internal Report of Nederlands Instituut voor Praeventieve Geneeskunde – TNO, 1970.
Dawson, G.D.	*A Summation Technique for the Detection of Small Evoked Potentials.* J. Electroenceph Clin. Neurophysiol. 6, pp.65-84, 1954.
Davis, R.C.	*Response Patterns.* Trans. New York. Acad. Sci. 19, pp.731-739, 1957.
Fenz, W.D. Jones, C.B.	*Individual Differences in Physiological Arousal and Performance in Sport Parachutists.* Psychosom Med. 34, pp.1-8, 1972.
Fukui, T. Morioka, T.	*The Blink Method as an Assessment of Fatigue.* Ergonomics. 14, pp.23-30, 1971.
Geddes, L.A. Baker, L.E.	*Principles of Applied Biomedical Instrumentation.* Wiley and Sons, New York, 1968.
Geddes, L.A. et al.	*The Impedance Pneumograph.* Aerospace Med. 33, pp.28-33, 1962.
Geddes, L.A. et al.	*Recording Respiration and the Electrocardiagram with Common Electrodes.* Aerospace Med. 33, pp.791-793, 1962.
Green, J.H.	*Blood Pressure 'Follower' for Continuous Pressure Recording in Man.* J. Physiol. 130, pp.37-38, 1955.
Halliday, A.M. Redfearn, J.W.T.	*An Analysis of the Frequencies of Finger Tremor in Healthy Subjects.* J. Physiol. 134, pp.600-611, 1956.
Hess, E.H.	*Attitude and Pupil Size.* Scientific American. pp.46-54, April 1965.

Mulder, G. Mulder-Hajonides van der Meulen, W.R.E.H.	*Mental Load and the Measurement of Heart Rate Variability.* Ergonomics. 16, pp.69-83, 1973.
Hyndmon, B.W. Gregory, J.R.	*Spectral Analysis of Sinus Arrhythmia During Mental Loading.* Ergonomics. 18, pp.255-270, 1975.
Kahneman, D. et al.	*Pupillary, Heart Rate and Skin Conductance Changes During a Mental Task.* J. Exp. Psychol. 79, pp.164-167, 1969.
Kate, F.H. Shannon, I.L.	*Identification and Significance of Parotid Fluid Cortico-steriods.* Acta Endocrinol. 46, pp.393-404, 1964.
Khalil, M. Otero, J.E.	*On the Quantification of Electromyography.* In: Proceedings 17th Annual Meeting Human Factors Society. pp.505-508, 1973.
Klinger, E. et al.	*Physiological Correlates of Mental Activity: Eye Movement, Alpha, and Heart Rate During Imagining, Suppression, Concentration, Search and Choice.* Psychophysiology. 10, pp.471-477. 1973.
Lader, M.H. Montagu, J.D.	*The Psycho-galvanic Reflex: A Pharmacological Study of the Peripheral Mechanism.* J. Neurol. Neurosurg Psychiat. 25, pp.126-133, 1962.
Lader, M.H.	*Pneumatic Plethysmography.* In: A Manual of Psychophysiological Methods. Venables, P.H. and Martin, Irene (Eds). North Holland, Amsterdam, 1967.
Lacey, J.I. Van Lehn, R.	*Differential Emphasis in Somatic Response to Stress.* Psychosom Med. 14, pp.71-81, 1952.
Lacey, J.I. Lacey, B.C.	*The Relationship of Resting Autonomic Activity to Motor Impulsivity.* Res. Publ. Ass. Res. Nerv. Ment. Diseases. 36, pp.144-209, 1958.
Littler, W.A. Honour, A.J. Sleight, P.	*Direct Arterial Pressure and Electrocardiagram During Motor Car Driving.* British Medical Journal. 2, pp.237-277, 1973.
Livia, A.R.	*The Nature of Human Conflict.* (Transl from Russian). Ed Gantt Liveright, New York, 1932. (Cited by Duffy. Reference 5).
Lowenstein, I.E. Loewenfeld, I.E.	*Electronic Pupillography.* Arch. Opthalmol. 59, pp.352-363, 1958.
Malmo, R.B. Shagass, C.	*Studies of Blood Pressure in Psychiatric Patients Under Stress.* Psychosom Med. 14, pp.82-93, 1952.
Marshall, J.	*Observations on Essential Tremor.* J. Neurol Neurosurg Psychiat. 25, pp.122-125, 1962.
Melton, C.E. Fiorica, V.	*Physiological Responses of Low-time Private Pilots to Cross-country Flying.* FAA Office of Aviation Medicine Report No.FAA AM-69-12, Washington DC, 1969.
Meyer, D.R. et al.	*Incentive, Anxiety, and the Human Blink Rate.* J. Exp. Psychol. 45, pp.183-187, 1953.
Monod, P.H.	*La Validite des Measures de Frequence Cardiaque en Ergonomie.* Ergonomics. 10, pp.485-537, 1967.
Montagu, J.D.	*The Psycho-galvanic Reflex: A Comparison of AC Skin Resistance and Skin Potential Changes.* J. Neurol Neurosurg Psychiat. 21, pp.119-128, 1958.
Mordkoff, A.M.	*The Relationship Between Psychological and Physiological Response to Stress.* Psychosom Med. 26, pp.135-150, 1964.
Plutchik, R.	*The Psychopsysiology of Skin Temperature: A Critical Review.* J. Gen. Psychol. 55, pp.249-268, 1956.
Rey, P. Rey, J.P.	*Effect of an Intermittent Light Stimulation of the Critical Fusion Frequency.* Ergonomics. 8, pp.173-180, 1965.

Rose, J.C. et al.	*Clinical and Investigative Application of a New Instrument for Continuous Recording of BP and HR.* New Eng. J. Med. 249, pp.615-617, 1953.
Sainsbury, P. et al.	*Symptoms of Anxiety and Tension and the Accompanying Psysiological Changes in the Muscular System.* J. Neurol Neurossurg. Psychiat. 17, pp.216-224, 1955.
Shaw, W.A.	*Facilitating Effects of Induced Tension Upon the Perception Span for Digits.* J. Exp. Psychol. 51, pp.113-117, 1956.
Simonson, E. Brozek, J.	*Flicker Fusion Frequency: Background and Application.* Physiological Review. 32, pp.349-374, 1952.
Smith, W.D.A	*A Method of Recording Systolic Blood Pressure.* British J. Anaesthesia. 34, pp.136-140, 1962.
Smith, W.S. O.Brien, C.	*A System for Rapid Analysis of Long-term Recordings of Heart Rate and Other Physiological Parameters.* Biomedical Engineering. 11, pp.128-131, 1976.
Stennett, R.G.	*The Relationship of Performance Level to Level of Arousal.* J. Exp. Psychol. 54, pp.54-61, 1957.
Sternbach, R.A.	*Principles of Psychophysiology.* Academic Press, New York, 1966.
Street, R. et al.	*The Evaluation of Mental Stress Through the Analysis of Parotid Fluid.* Human Factors. 15, pp.543-555, 1970.
Thackray, R.I. et al.	*Personality and Physiological Correlates of Performance Decrement on a Monotonous Task Requiring Sustained Attention.* British J. Psychol. 65, pp.351-358, 1974.
Tursky, B. et al.	*Differential Patterns of Heart Rate and Skin Resistance During a Digit Transformation Task.* J. Exp. Psychol. 83, pp.451-457, 1970.
Venables, P.H. Martin, Irene (Eds)	*A Manual of Psychophysiological Methods.* North-Holland, Amsterdam, 1967.
Voskresenskiy, A.D. Venttsel, M.D.	*Statistical Analysis of the Cardiac Rhythm and Haemodynamic Indices in Physiological Research.* Problems of Space Biology. No.1. 26, Technical Translation No.F-16196, NASA, Washington DC, 1975.
Weinman, J. Manoach, M.	*A Photoelectric Approach to the Study of Peripheral Circulation.* American Heart J. 63, pp.219-231, 1962.
Wenger, M.A.	*Pattern Analyses of Automic Variables During Rest.* Psychosom Med. 19, pp.240-244, 1957.
Wenger, M.A. et al.	*Autonomic Response Specificity.* Psychosom Med. 23, pp.185-193, 1961.
Wennberg, A. Zetterberg, L.H.	*Application of a Computer-based Model for EEG Analysis.* Electroenceph Clin. Neurophysiol. 31, pp.457-468, 1971.
Whitney, R.J.	*The Measurement of Changes in Human Limb-volume by Means of a Mercury-in-rubber Strain Gauge.* J. Physiol. 109, pp.5-6, 1949.
Wilkinson, R.T.	*Muscle Tension During Mental Work Under Sleep Deprivation.* J. Exp. Psychol. 64, pp.565-571, 1962.
Young, F.A. Biersdorf, W.R.	*An Apparatus for Taking Rapid Photographs of the Pupil in Visual Darkness.* American J. Psychol. 65, pp.617-618, 1952.

CHAPTER 4

OBJECTIVE METHODS

by

W.Dean Chiles
Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma
USA 73125

CHAPTER 4

OBJECTIVE METHODS

by

W.Dean Chiles
Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma
USA 73125

CONTENTS

OBJECTIVE METHODS

4.1 INTRODUCTION

The concept of workload is of special interest in that there is abundant evidence (at least of an anecdotal nature) that workload can be a "go/no go" modifier of the performance of the pilot as a functional subsystem, especially under emergency conditions. Therefore, finding or developing an appropriate methodology that yields *reliable* and *valid* measures of pilot workload is a goal that, if achieved, should lead to important gains in safety and mission accomplishment through the resultant system design and procedural modifications.

Our ultimate concern in the measurement of workload must be the determination of the manner and extent that workload affects the probability of mission success. Thus, in this context, it is appropriate to raise the traditional engineering questions related to the probability of "failure" of the pilot as a functional subsystem. From the point of view of reliability engineering, we might say as a first approximation that an *acceptable* level of workload for a given phase of a mission would be characterized by a set of system-induced (*system* in its broadest sense) task demands such that the probability is equal to or greater than some specified value that the pilot will be able to satisfy those demands and successfully complete that mission phase without compromising subsequent mission phases. (Clearly, the probability value selected for one-time, high-priority missions, for multiple missions, and for routine operations would likely be different.)

The literature in this area is quite clear on one point. There is no generally accepted definition of the term "workload". Some authors would use the term primarily to refer to input loading; e.g., the number and nature of the displays (and controls) that must be used by the pilot in performing his job. Others would use the term to refer to how hard the pilot has to work; these authors tend to prefer biomedical and/or subjective indices of workload. Still other authors emphasize those aspects of workload that relate to performance; e.g., speed and accuracy of response.

4.1.1 A Working Definition of Workload

No attempt will be made to arrive at a formal, comprehensive definition of workload; the problems in developing such a definition are numerous and formidable (see Chapter 1). However, it seems necessary to offer some sort of working definition – even though it be rather nonspecific and largely descriptive – of the way the term will be used here before meaningful discussion of measurement methodology in the area can be undertaken. Therefore, for the purposes of this chapter, *level of pilot workload* will be assumed to be a hypothetical concept that is determined by or related to the aggregate of the task demands placed on the pilot by the system during some relatively short-duration mission or phase of a mission coupled with the actions required of the pilot to satisfy those task demands. The actions required may be overt or they may be covert. They may be physical, they may be mental, they may be perceptual, they may be oral, or they may be some combination of any or all of these. There may be purposes for which it is appropriate to talk about system demands independent of pilot actions in considering workload. However, in the present discourse it will be assumed that, to the extent a system demand is not followed by suitable and timely action on the part of the pilot, the mission phase will have been completed in less than an acceptable manner (if it is completed at all). In other words, demands that do not require action (either overt or covert) are not *really* demands; and actions that are initiated for reasons other than to satisfy a system demand (and are potentially disruptive of mission accomplishment) should be eliminated by training and operating procedures. Thus, "stimulus" and "response" will not be treated separately.

Although for purposes of exposition a general definition of workload is adopted in this chapter, it should be clearly understood that the goals and intents of a given measurement study are the important determiners of how workload should be defined and what methodology should be adopted for a specific application. For example, one designer/researcher may need to know simply which of two alternative – but otherwise satisfactory – single-purpose displays makes a smaller contribution to the pilot's workload. Another designer/researcher may need to know how quickly, if at all, the pilot can manually operate a device that is normally hydraulically or electrically powered. Numerous other differences in purposes and, hence – by implication – methodologies can be readily imagined. More will be said on this topic later, but it is not our intent to be dogmatic – especially about unsettled issues.

4.1.2 Chapter Outline

The remainder of this chapter will consist of six sections: Some Rudiments of Measurement Theory; Laboratory Methods; Analytic and Synthetic Methods; Simulation Methods; In-Flight Methods; and Discussion, Recommendations, Cautions, and Conclusions. The approach that will be used in the research-oriented sections will be to describe selected programs in which particular methodologies have been applied, and, where appropriate, data will be presented to give an indication of the kinds of results achieved. No attempt at a comprehensive review will be made; the reader is directed to companion chapters and to a number of suitable References 1, 2, 3, 4, 5.

4.2 SOME RUDIMENTS OF MEASUREMENT THEORY

This section is not in any way intended to be a definitive exposition on measurement theory. However, certain basic concepts of measurement theory will come up in later sections and it seems expedient to mention and briefly explain them before proceeding. (Some readers may wish to skip this section.)

4.2.1 Validity

The first and perhaps most important notion to be dealt with is validity. Ultimately, this simply means, "Are we really measuring what we intend to be measuring?" The answer to this question, in the most precise use of the term, assumes the existence of a criterion. For example, in the field of selection, we might want to select only those aviation candidates who have a high probability of completing flight training; our criterion, then, would be successful completion of training (and perhaps final average grade). The validity of the selection measure would thus be determined by the accuracy with which it predicts which trainees will graduate. Unfortunately, in the workload areas we have no such criteria, and, therefore, we must rely primarily on what is called "content validity" – which really amounts to expert, professional opinion. Still another kind of validity, "face validity", can be important in motivating test subjects; in this sense, (face) validity means the test situation *appears* to be like the job of the pilot. (No small part of the expense of building simulators is devoted to trying to achieve face validity.)

4.2.2 Reliability

Reliability has several meanings that are applicable in varying degrees to the problem of workload measurement. In one use, it refers to the engineering characteristics of the measurement system and relates to the repeatability of a measure or phenomenon; with a constant known input, what is the variability of the output? That is, how accurately can the output be predicted from the input? Reliability in this sense involves internal characteristics of the test device, and the term is used to reflect the sensitivity of a measurement procedure to, for example, temperature changes, drift characteristics of components, etc. A second, closely related use of the term "reliability" depends not only on the above characteristics of the test equipment but also on the human behaviour being measured. For example, in even the most carefully controlled experimental situation, the response latency of the human subject to the onset of a light will show variation across trials and across individuals; the amount of such variation will depend on the behavior being measured. In this use of the term, an approximation of the reliability estimate can be obtained by observing the extent to which a group of individuals shows the same rank ordering on each of two measurements of the phenomenon per individual. This is generally referred to as test-retest reliability. It should be noted that the apparent reliability (i.e., the size of the reliability coefficient) is dependent on both the true reliability of the test or equipment used and the existence of *stable individual differences in the behavior being measured*. Thus, with highly trained, highly selected, skilled operators, the variability for a given individual from trial to trial may be as great as the variability across individuals on a given trial. Under such conditions, the measured reliability could appear to be rather low even though the basic measures are quite stable. In any case, if meaningful comparisons are to be made concerning workload variations, some estimate of the stability and precision of the measures must be secured. Otherwise, there is no way to determine whether an obtained difference in a measure is properly interpreted as being real or as being a result of chance factors.

4.2.3 Sensitivity

In any evaluation of alternative system designs or system operating procedures, it is necessary to have some index of the sensitivity of the measures to the variables being manipulated. For example, simple reaction time to an attention-getting signal calling for a single response is quite stable even when there are large changes in presumably important variables. The same is true of many simple tracking tasks. Perhaps the main reason for this stability is the extreme adaptability of the human operator. If the operator is confronted with a task situation in which he can concentrate all of his resources on the performance of the task, then, at least for relatively short intervals, he can maintain his performance of single tasks amazingly well. Thus, for example, if altitude were a variable of interest and simple reaction time were the measure used, we would conclude that performance is not impaired until the pressure altitude is somewhat in excess of 5,000 meters. Thus, such simplistic approaches could lead to questionable conclusions. What all of this means is that it is sometimes necessary either to do preliminary research or to add variables to the main research simply to get an index of the sensitivity of the measurement procedure to relevant variables.

4.2.4 Magnitude of Effect

If two alternatives (displays, for example) are exactly equivalent in terms to cost, weight, size, etc., then any reliable (statistically significant) superiority of one alternative over the other is sufficient basis for choosing the better alternative. However, if there are important differences between the two in terms of cost, weight, etc., then it is necessary to establish not just the statistical significance of a difference (if there is one) but, especially if the more expensive one is the better, how much better it must be to make in fact a *practical* difference. Expert, professional judgment plays a major role here.

4.3 LABORATORY METHODS

From the point of view of methodology, there are several characteristics of "laboratory" methods that make them highly desirable. First, for most laboratory tasks, it is possible to exercise very precise control over the performance demands imposed on the operator. One can with relative ease control the number of tasks that are active, the rates at which signals are presented, and the timing of the signals on individual signal sources as well as across sources. Second, "exact" duplication of test procedures is readily achieved. Third, laboratory methods in general can provide the highest precision of measurement that one is likely to achieve in the realm of operator behavior. Fourth, depending on the level of complexity of the experimental task structure, high equipment reliability is possible at relatively modest costs, and, because physical safety is not involved, any lack of mechanical or electrical reliability is primarily just a source of inconvenience. In addition, tasks can be selected and structured so that good test-reset reliabilities are common. And fifth, it is generally not terribly difficult to establish the sensitivity of the task measures to variables of known operational importance and behavioral potency.

4.3.1 Laboratory Methods

Early in the history of behavioral sciences, there was considerable interest in the area of mental load in what would now be called an information processing context. These early efforts were directed at an attempt to break down complex reaction time into its constituent components. To illustrate how this breakdown was approached, assume that the operator is confronted with a red light on the right of a display and a green light a few centimeters to its left. Assume further that two response buttons are conveniently located for the use of the right hand. The subject is instructed to depress the rightmost button if the red light comes on and the left button if the green light comes on. Thus, the subject must decide which light came on and which button is correct. Assume now a different procedure: a number of responses are recorded in which only the red light and the rightmost button are present and other responses when only the green light and the leftmost button are present. With this procedure, the subject only has to become aware that a light is on and respond. The notation is that the difference between the average response time to the single-light/single-button conditions and two-light/two-button condition provides an estimate of the "mental" processing time in recognizing whether the red or the green light has been illuminated in the latter condition. This general procedure has been expanded and permutated in a variety of ways. The well-established result is that if N signals are uniquely coordinated to N possible responses, then:

$$\text{Reaction Time} = a + b \log_2 N$$

where a and b are constants.

Thus, it is seen in this very elementary case that performance is a function of task demand or workload.

4.3.2 Timing – Speed and Load Stress

Another line of laboratory research has been concerned with the timing of response in a monitoring situation. The notion of timing in skilled performance was first introduced by Sir Fredrick Bartlett[6]. The concept was further refined by Conrad[7], who proposed to define timing (of responses) as "creating the most favorable temporal conditions for response". Conrad treated *load* in his studies as being a function of the number of signal sources and considered *load stress* to be produced by increasing that number beyond some value. He used the term *speed stress* to refer to excessive rates of presentation of signals from a given source (or number of sources). Conrad found that subjects tended to alter the point of response initiation in a manner apparently designed to even out, temporally, the sequence in which they were required to take action. In a later study, Conrad[8] gave subjects limited control over the average rate at which signals would appear; this control gave subjects the opportunity to slow down the signal rate so they could successfully respond to essentially concurrent signals on separate displays; on the average, subjects did better under this condition. These results suggest the desirability of, wherever possible, adopting designs and operating procedures that permit latitude in the exact point at which events must be initiated by aircrew personnel.

Knowles, Garvey, and Newlin[9] investigated speed and load effects in a different context; they were interested in display-control compatibility relationships. The part of their experiment that is of particular interest here is the comparison of a 10 x 10 matrix of lights (associated with a 10 x 10 matrix of response buttons) and a 5 x 5 matrix of lights (associated with a 5 x 5 matrix of buttons). The rate of presentation of information (not signals) was equalized across the two conditions; the rates used were 1.75, 2.25, 2.75, and 3.0 bits/second. They found that the effect of load (display size) had a greater effect on error rate than did rate of presentation of signals. (See Table 1 overleaf.) They also found, incidentally, that subjects could respond at an average rate of 0.45 signal per second without errors in a self-paced mode whereas when the task was forced-paced at that same rate, subjects made 36 percent errors.

4.3.3 Secondary Loading Tasks

One general, more direct approach to the study of workload in the laboratory has been through the use of secondary or loading tasks. Knowles[10] summarizes early work of this sort and provides the general rationale for the application of the technique to workload measurement in a part-task simulation context. Knowles (page 156) states that auxiliary tasks are used " . . . with the intention of finding out how much additional work the operator can undertake while still performing the primary task to meet system criteria".

TABLE 1

Mean Errors per 100 Stimuli*
Speed (bits/s)

Matrix	1.75	2.25	2.75	3.0
Small (5 x 5)	2.5	3.6	4.1	7.1
Large (10 x 10)	3.6	10.8	13.1	15.8

* Adapted from Knowles, Garvey, and Newlin[9].

"Secondary tasks are used because primary part-task performance measures, in and of themselves, seldom reflect operator-load. . . . they seldom tell the price paid in operator-effort in meeting (the system) criterion." Knowles goes on to describe an earlier study, Knowles and Rose[11], in which a simulated lunar landing task was being investigated. He says that in that study: "The loading scores were sensitive to differences in problem difficulty; they reflected increased ease in handling the control task as a function of practice; they revealed differences in workload between members of a two-man crew; and they showed that the particular control law under consideration was unsatisfactory because of the extreme buildup of operator load during the last few seconds of the landing. *None of these results was available from system performance criteria*; i.e., time, fuel, miss-distances." (Emphasis added.) The basic approach in this method is to compare the levels of performance achieved on the "loading" task when performed alone with the levels achieved when it is performed in combination with the primary task; this difference is said to provide an index of the workload imposed by the primary task.

Benson, Huddleston, and Rolfe[12] reported a study in which, among other things, they evaluated a one-dimensional tracking task by using two altitude displays; performance was measured with each display with and without a secondary light-acknowledgement task. They found a small, consistent superiority of a counter-pointer display over a counter-only display with the tracking-only condition. When the secondary task was added, they found significant decrements in tracking with both displays with a significant superiority of the counter-pointer over the counter-only display. The secondary task showed significant decrements when added to either tracking task; the differences between display conditions were fully compatible with the findings for the tracking task – namely, the display that showed the better performance of the secondary task. They interpret the decrements in the primary tracking task to pose serious questions as to "the essential feature of the subsidiary task situation; namely, that consistent primary task performance is possible in two task conditions". Benson et al.[12] instructed their subjects that they were to attend to the secondary task only when they could properly do both jobs together. They interpret their results to suggest that subjects may not be able to comply with such instructions and discuss at some length whether and how subjects might be able to perceive that their performance is being maintained on the primary task. They also suggest the possibility that a continuous primary task may be more likely to suffer decrements than a discrete primary task. Depending on the frequency characteristics of the display disturbances and the time it take the subject to perceive which light has been illuminated, it is quite reasonable to expect that, on a probabilistic basis, looking at and responding to their secondary task would encourage error accumulation on their primary task.

It should be noted that Benson et al.[12] concluded that "there is no doubt that the presence of a second task added to the value of the experiment . . . .". Thus, their discussion of the changes in the primary task is related primarily to "theoretical" expectations as to how the secondary task technique should operate in practice. It could be argued that their experiment actually demonstrated two important findings: (1) the counter-pointer display is better in that it resulted in better performance (numerically in the case of tracking only and statistically in the case of the two-task situation); and (2) the counter-only display is more sensitive to possible distraction or interference from other tasks.

The question can also be raised as to whether the subsidiary task technique *necessarily* relies on the subject's achieving parity of performance on the primary task between the one- and two-task conditions. Clearly, Benson *et al.*[12] demonstrated in their experiment that useful information can be obtained from the technique when this assumed state of affairs does not obtain. If we consider one of the empirically based reasons that Knowles pointed at in using the technique, it is frequently the apparent absence of an effect on single tasks of possibly important variables that suggests the possible value of using secondary operator loading tasks. Thus, it could be argued that so long as changes in the primary task and the secondary task are compatible (i.e., lead to the same conclusions), we should not be overly concerned about changes in the primary task – changes that may be *valuable data* in and of themselves.

Senders (Reference 13, p.208) says there are four assumptions that underly the secondary loading task methodology: (1) The operator is a single-channel system. (2) The channel has a fixed capacity. (3) the capacity has a single metric by which any task can be measured. And (4) the constituents of workload are additive linearly, regardless of the sources of the load. These assumptions are required if *channel capacity* is to be given formal status as that term is used in information theory. However, in the practical application of the secondary loading task methodology, it is suggested that the first and second assumptions stated by Senders are of major significance only under certain conditions – for example, when neither the primary task performance nor the loading task performance changes when the two are performed simultaneously. In that event, although we would have learned something interesting about the two tasks, we could not be sure whether the primary task represents a "no load" condition, the operator has employed a previously "unused" channel, the operator has simply "expanded" his (single) channel capacity, or, what is most likely, the time requirements of the two tasks are such that the performance of neither interferes with that of the other. The possible absence of linear additivity places a heavy burden of responsibility on the choice of the loading task; clearly, the loading task must have properties in the "additivity domain" that warrant generalization to the kinds of system tasks that might be coupled with the primary task being investigated. By the same token, the metric implied by the secondary task must also be applicable to possible system task requirements.

Perhaps the safest interpretation of the changes in the secondary task would be that they serve as an index of the *spare time* that the operator has while performing the primary task at criterion levels. But even in this interpretation it is necessary to make some kind of assumption regarding the ease of back-and-forth transition (primarily in terms of time) between the primary task and the particular secondary task being used. Rolfe[14], who provides an excellent review and discussion of the secondary task method of measuring workload, closes with the following caution: "The final word, however, must be that the secondary task is no substitute for competent and comprehensive measurement of primary task performance. The technique should always be looked upon as a means of gathering additional information rather than an easy way of gathering primary information." This caution should not be taken lightly, even though the study of Knowles and Rose[11] showed secondary task measured to be sensitive to important factors not revealed by the primary task measures.

4.3.4 Cross-Adaptive Loading Tasks

Kelley and Wargo[15] take the position that consistent performance on the primary task is vital. They offer data from a demonstration experiment using two subjects in which decrements on primary and secondary tasks are apparently not compatible; conditions that were ranked, in order of merit, A, B, C on the primary task were ranked B, A, C by measures from a secondary task. Their primary task was a two-dimensional, two-display compensatory acceleration tracking task; the secondary task consisted of two identical "warning" lights, one above the other, located where subjects could see them by peripheral vision but had to look at them directly to determine which light had been illuminated; response to the lights was made with a thumb switch located on the tracking control stick. When the lights task was active, one of the lights, selected at random, would turn on 0.44 second after the subject extinguished the previous light. The primary task variable of interest was display gain, of which there were three levels. Three test conditions were used: primary task only, primary task plus the loading task with independent programming (straight subject pacing), and primary task plus "cross adaptive" programming of the loading task. In this latter case, as long as tracking error (vector root-mean-square (RMS)) remained below the criterion level, one of the lights would be turned on as noted above. If error exceeded the criterion level, the lights task would be deactivated until tracking error again was below criterion. It is important to note that Kelley and Wargo[15] instructed their subjects to perform both tasks ". . . as well as they could and not to neglect one for the other". Thus, the concepts of *primary* and *secondary* are somewhat blurred; the experimenter, without informing the subjects, had arbitrarily decided which was which. The previously mentioned findings from Kelley and Wargo, in which the inferences from the primary and secondary task performances were not compatible, were taken from the condition involving tracking plus the subject-paced loading task. The compellingness of their results suffers from several problems. First, only two subjects were used. Second, the display gain variable was significant for the tracking-only condition. Third, the display gain variable was significant for the subject-paced loading-task condition for one subject though not for the other. And, fourth, a cleaner evaluation of the cross-adaptive approach to using loading tasks would have resulted if task priorities had been clearly specified.

However, the approach, overall, looks interesting and further evaluation of its characteristics *vis-a-vis* traditional loading-task procedures would appear to be warranted.

4.3.5 Memory Scanning Tasks

Another variation on the secondary task technique has been described by O'Donnell[16]. This procedure is "an adaptation of an item recognition technique first described by Sternberg"[17,18]. The basic approach is that the operator is required to learn a set of positive stimuli (so-called because their appearance calls for a positive response). Members of the positive set, frequently letters of the alphabet, are presented one at a time; generally, on half of the trials the stimulus is a member of a negative set. On the appearance of a letter, the operator is instructed to respond as quickly as possible by depressing a "yes" key if the letter is a member of the positive set and a "no" key if it is a member of the negative set. Under appropriate conditions, a linear relation exists between the size of the positive set (typically 1 to 8) and reaction time The psychological theory behind the use of this task is that average reaction time with a given number of stimuli in the positive set can be broken down into three parts: (1) stimulus encoding, (2) memory scan, and (3) response selection and execution. For a given set of conditions, the first and third parts are assumed to be constant, whereas the second part is interpreted to be a direct reflection of memory scan speed and/or memory load. Thus, changes in the y-intercept value (i.e., the response time for the primary task alone) are assumed to reflect changes in the perceptual and/or response aspects of the task. Changes in the slope of the curve are assumed to reflect changes in the rate at which memory is scanned and/or the amount of memory load involved. In other words, the y-intercept value serves the same function as a measure from a secondary loading task as described previously; the higher the intercept (i.e., the longer the average response time), the greater the assumed loading produced by the primary task. In addition, a change in the slope of the response-time curve *might* be interpretable as a reflection of the amount of memory load added by the primary task. The value of this task as a loading task in the usual sense has been borne out by the results of preliminary studies conducted thus far. However, the possibilities with respect to its providing a measure of *memory load* are still to be demonstrated. It should be noted that earlier results reported by Darley, Klatzky, and Atkinson[19] suggest that *the addition of memory load not directly related to the item recognition task does not affect the slope of the reaction time curve*.

4.3.6 Synthetic Work Tasks

Operator workload has also received attention in an area of laboratory research that is concerned with "synthetic work". The rationale for the development of synthetic work tasks has been described in detail elsewhere (Chiles, Alluisi, and Adams[20], and Chiles[21]); however, for those readers to whom the notion is new, a brief description of the techniques and philosophy will be given here.

The point of departure of the synthetic work approach is a *behavioral* analysis of the performance requirements placed on the operator by some particular aviation system or by a class of such systems in general. Tasks are then selected against a criterion of content validity (i.e. tasks are selected because they measure functions judged by experts in the field to be important to aircrew operations) as well as a general criterion of face validity (i.e., the tasks are configured to be acceptable to target populations, such as pilots). Consumer acceptance of the tasks has always been good[20]. The resultant hardware is designed so that the selected tasks can be presented in any combination desired and individual tasks can be varied along both time constraint and task difficulty parameters. The original goals of the program in which the particular system to be described was conceived were the evaluation of procedural (e.g., work schedules), environmental (e.g altitude), and pharmacological (e.g., alcohol) variables as these factors might affect complex performance.

Within the context of the way these tasks were developed and have been used, the notion of workload is a relative concept. However, from the beginning it was assumed that it would be desirable, if not necessary, to vary the apparent workload imposed on the operator from very light to near overload; overload is defined, for this purpose, as decrements on all or most of the concurrently performed tasks, even in the absence of any external stressor. Thus, extensive data have been collected on a variety of task combinations that, on a rationally defensible basis, would be expected to correspond to different workloads.

The specific tasks used involve monitoring of lights and meters (providing measures of reaction time), mental arithmetic, pattern discrimination, elementary problem solving, and two-dimensional compensatory tracking. The task combinations used in a study by Hall, Passey, and Meighan[22], involving an earlier version of what is called the Multiple Task Performance Battery[20], are shown in Table 2. Note that two basic conditions were examined – monitoring tasks only

TABLE 2

Performance Schedule*

	Monitoring Only								*Complex*							
Auditory Vigilance	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Warning Lights	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Meter Monitoring	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Mental Arithmetic										X	X					
Problem Solving (Group)											X	X	X	X		
Pattern Discrim.														X	X	
15-Minute Interval	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8

* Adapted from Hall, Passey, and Meighan[22].

TABLE 3

One-Hour Task Schedule

Warning Lights	X	X	X	X
Meter Monitoring	X	X	X	X
Mental Arithmetic	X	X		
Tracking, Two-Dimensional	X			X
Problem Solving (Individual)		X	X	
Pattern Discrim.			X	X
15-Minute Interval	1	2	3	4

and "full battery" as specified in Table 2. If it is assumed that the subjects tended to treat the monitoring tasks as secondary (loading) tasks, then the performance levels on those tasks can be considered to be an index of the workload imposed on the operator by the different combinations of the other tasks. Figure 1 shows the response latencies on a normalized scale for the responses to the offset of any one of five green lights located one at each corner and one in the middle of the test panel. Figure 2 shows response times in seconds for the detection of a shift in the average value of the "randomly" wandering pointer of any one of four meters located across the top of the test panel. Each of these figures contains two curves – one for the given monitoring task performed with only the monitoring tasks active and one for monitoring performance as a function of the different "active task" combinations. Note that the first and the last points of the curves labeled "full battery" consist of only the monitoring tasks, thus providing "anchor points" for the curves. The normalizing scale applied to the data for the green-lights monitoring tends to suppress the apparent amplitude of the shift in response times, but the changes across task combinations are statistically significant. The changes in the meter-monitoring task are much larger and, of course, are also statistically significant.

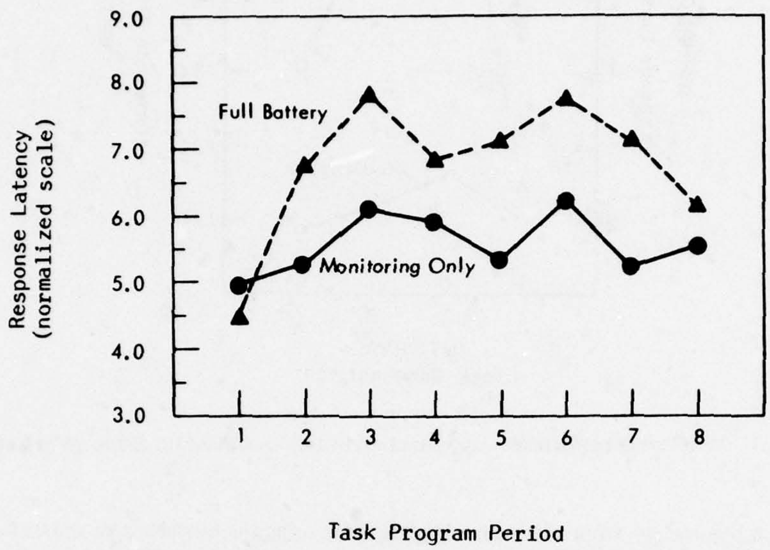

Fig.1 Mean response latency in detecting green warning-light signals during each 15-minute period of the basic 2-hour task program. (Adapted from Hall, Passey, and Meighan[22])

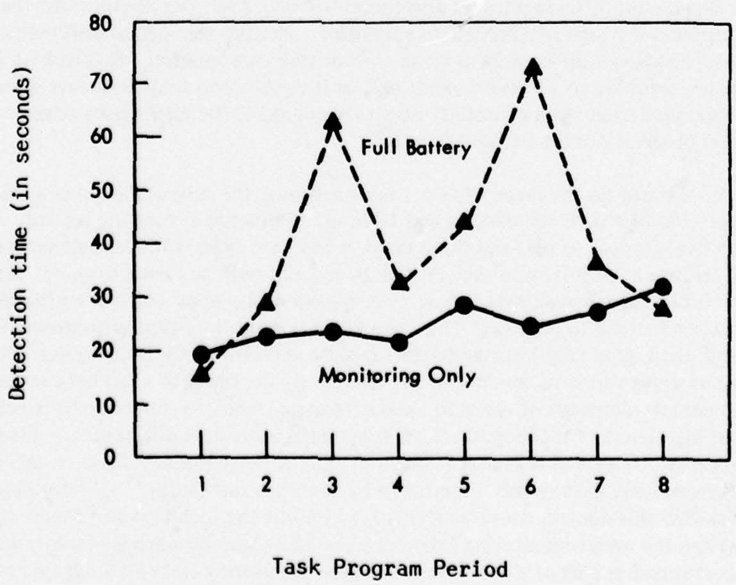

Fig.2 Mean detection time for correct detections of probability monitoring signals during each 15-minute period of the basic 2-hour task program. (See Table 2.)

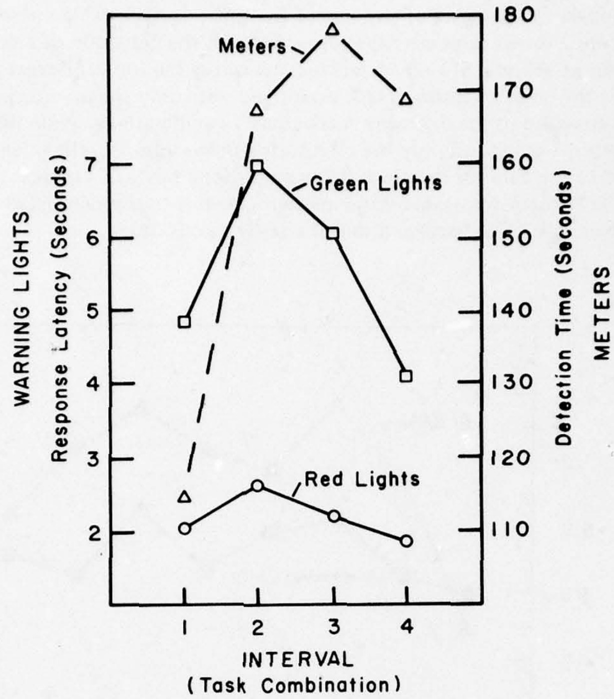

Fig.3 Monitoring performance as a function of task combination as shown in Table 3

The data shown in Figure 3 are from a later unpublished study using the task schedule shown in Table 3 and using pilots as the subjects. Figure 3 shows response times in seconds to the onset of red lights (physically paired with the green lights) and the offset of green lights. Figure 3 also shows the detection times in seconds for the meter-monitoring task. (Although the tasks are functionally the same as those used by Hall et al.[22], the data of these two figures were collected by using a new, computerized version of the Multiple Task Performance Battery.) For all three task measures, the differences across task combinations are significant. (It may or may not be important that the longest response times for the light-monitoring tasks were associated with a different task combination than were the longest response times for the meter-monitoring task.) Significant differences were also found between task combinations for the tracking task (vector RMS error) and for the problem solving task (redundant responses). Neither the mental arithmetic task nor the pattern discrimination task showed significant differences as a function of task combination. This lack of differences could mean that these latter tasks are less sensitive to workload variations, or it could mean that they were given higher priorities by the subjects. Although a detailed evaluation of exactly how to account for the differences across tasks is not relevant to our purposes, some general observations are perhaps in order.

The data of Figure 3 are based on the mean of two 1-hour sessions; the subjects had had a total of about 7 hours of practice on the tasks before the first of these sessions and 10 hours of practice before the second. Among the literally hundreds of subjects who have learned to perform these tasks, it has been typical that the subjects initially have difficulty, for example, completing arithmetic problems in the allotted 20 seconds with any time to spare. Similarly, they frequently get "hung up" on the problem-solving task at the expense of the other tasks, even though they are reminded during training that they are to attend to all tasks. Thus, the learning procedure typically consists of first, acquiring skill on the individual tasks and, then, gradually learning to shift rapidly and efficiently from a given active task on which their attention may be focused at a given time to concurrent demands (e.g., the onset of a red light or another active task); or, on satisfaction of the momentary demands of the active tasks, they may shift to scanning the panel for monitoring signals. It is also clear that even at high levels of training, there are substantial individual differences in the smoothness and speed with which attention appears to be shifted from exercising one kind of behavioural process to another, different kind of process. For this and other reasons, a study was undertaken by Jennings and Chiles,[23] to determine whether an independent (time sharing?) skill in this domain could be identified by using the techniques of factor analysis. In this study, the lights (red and green) and the meter-monitoring tasks were found to load on separate factors when performed as individual tasks. When performed as part of a complex task, these monitoring tasks all loaded on a third, independent factor. If these results, which suggest a possible time-sharing ability, should hold up on replication, important implications are suggested for the selection of subjects to be used in various kinds of tests of systems and system components.

The synthetic work methodology has yielded other results of relevance to the use of secondary loading tasks as measures of workload. In a study of the effects of blood alcohol levels of approximately 0.1 percent, a device that was different from the Multiple Task Performance Battery described above was used, but the requirements for time sharing were similar; performance of different combinations of mental arithmetic, monitoring, and two-dimensional tracking tasks was required[24]. The results showed that the monitoring tasks were affected at each of the two levels of workload used, but the tracking task was affected only at the higher of the two workloads (tracking, monitoring, and arithmetic). The arithmetic task was not significantly affected under either workload condition. In this study, the subjects apparently regarded the arithmetic task as being a "primary" task and gave it priority over the other tasks; it could perhaps be argued that the subjects "protected" their arithmetic performance at the expense of the other tasks. When just the tracking and monitoring tasks were presented, it could similarly be argued that they placed priority on the tracking task and "protected" that performance. Whether or not these proposed interpretations are accepted as reasonable, it seems clear (and commonsense) that the priority an operator assigns to a task will be an important factor in determining the level of performance maintained on that task as other duties are added.

4.4 ANALYTIC AND SYNTHETIC METHODS

The methods to be discussed in this section have been somewhat arbitrarily categorized as analytic or synthetic. (Both types of methods have some elements of each general approach, but the first to be discussed probably leans a little more in the analytic direction and the second, a little more in the synthetic direction.)

4.4.1 Analytic Method

Senders has been a major proponent of the analytic method of workload analysis[25,26,27,28,29]. This basic approach rests on the following assumptions listed by Senders, (Reference 13, p.209):

(1) *Visual* distribution of attention is the *major* indicator of operator workload.

(2) The various signals that must be monitored demand *attention* commensurate with the characteristics of the signal and the required precision of readout of the signal by the human operator.

(3) The human operator is effectively a *single-channel* device capable of attending to only one signal at any time.

(4) The probability of human *failure* at any time is equal to the probability that *two or more* signals will demand simultaneous attention.

Senders states that these are simplistic assumptions in the sense that other signal sources (e.g., auditory) are not considered; attention to the visual part of continuous manual control tasks is not considered; and peripheral vision is not taken into account. Thus, the major analyses have to do primarily with instrument layout and deal only with requirements for instrument reading as a source of workload.

An important feature of this approach is that it can be applied in advance of the existence of specific hardware; it requires only that certain conditions be specifiable. For a given visual display, if the following information is available, then workload-related parameters can be calculated:

(1) The maximum or cutoff frequency of the display must be specified. From this figure, the required fixation frequency as a function of time can be calculated.

(2) Signal amplitude and acceptable error of reading must be specified.

From (1) and (2) the information rate for the display can be calculated. From the information rate, the fixation duration can be calculated (on the basis of the known relation between information content and response time). The product of fixation frequency and duration of observation yields the time required for observing the display expressed as seconds/second. The times found for each display instrument can be summed to get an index of monitoring workload as total seconds/second required overall in observing instruments. If uncorrelated signal sources are assumed, transition probabilities (e.g., probability of looking at display B after having observed display A) can be calculated and thus lead to guidelines for optimum instrument layout.

Sender[27] tested these notions in a laboratory situation by using four meters that were driven at different frequencies. He then compared predicted fixation frequencies based on the display characteristics with fixation frequencies as determined by motion pictures of the eye positions of the subjects. The agreement between prediction and data was quite good. Subsequently, Carbonell, Ward, and Senders[30] compared predictions with data from pilots flying approaches to landing in a simulator. Instrument pickoffs were used to establish the frequency characteristics of the various instrument displays and eye-movement measures were used to determine fixation frequencies. The agreement between the values from the prediction procedures (Nyquist model) previously used,[13] and the data was reasonably good; however, a *queueing theory* model gave substantially better agreement.

Clement, Jex, and Graham[31] describe the application of a "manual control-display theory" to instrument landings of a "large subsonic jet transport". This theory, detailed by McRuer and Jex[32] and McRuer, Jex, Clement and Graham[33], attempts to use hypothesized ratios between fixation frequencies and display bandwidths that are tailored to the accuracy-of-control requirements for the particular display. Then, using a procedure otherwise similar to that described by Senders[13], Clement et al.[31] computed a fractional scanning workload index for each display function and summed these arithmetically to get a quantity that is equivalent to a seconds/second scanning index. They showed that, as a design exercise, the predicted scanning workload for a selected aircraft panel layout could be reduced from 1.32 (anything greater than 1.0 is overload) to 1.01 by combining certain displays. Although their predicted best display arrangement "agrees with that actually adopted" by a major airline for FAA Category II certification, empirical validations of scan times and fixation durations are not presented. In a subsequent study, Weir and Klein[34] collected data by using a "DC-8" flight simulator; however, their results in terms of scan times were compared with previous findings with aircraft and simulators rather than with theoretical predictions based on display information. Further discussion of this analytic approach can be found in Allen, Clement, and Jex[35].

The analytic approach to workload prediction requires considerable knowledge about the characteristics of the forcing functions of the various instruments and displays. But, where such information is available, the methodology developed to date shows promise, especially in applications to new, design-stage systems. However, substantial effort in the empirical validation of the procedures is still needed and warranted.

4.4.2 Synthetic Method

What is being referred to here as the synthetic method might equally well be called a *combinatorial* method. The point of departure of this method is a task analysis of the system; the proposed mission or operating profile is broken down into segments or phases that are relatively homogeneous with respect to the way the system is expected to operate. For each such mission phase, the specific performance demands placed on the operator are identified through task analysis procedures. Once individual tasks and subtasks have been isolated, previously available (e.g., Munger, Smith, and Payne[36]) or *ad hoc* data are compiled on the performance of the tasks with both performance times and operator reliabilities being taken into account. The information on performance times is then accumulated for a given mission phase and the resultant sum is compared with the predicted duration of the phase. The comparison of these two quantities – time required to perform versus time available – can be used to reflect an index of workload. Although other factors can be included in this synthesizing process, time is typically the primary variable considered.

One example of this approach is the Cockpit Evaluation and Design Analysis System described by Brown, Stone, and Pearce[37]. Brown et al. define workload as follows: "Flight crew workload is the ratio of the summation of required crew-equipment performance time to the time available within the constraints regulated by a given flight or mission". Their design and analysis system is computerized and is organized in such a way that detailed information can be included regarding required times, available times, items of equipment involved, and flight phases as well as the design personnel responsible for the various equipments and subsystems.

Flight phases are further broken down by identification of what they call milestones, a milestone being a change in heading, airspeed, altitude, etc. Preliminary allocations of duties and activities are based on operating techniques of expert pilots and operating procedures for similar aircraft. For purposes of workload prediction for a given segment, the computer output is expressed in the form of percentage-of-capacity figures for each task element each crew member is to perform. In this way critical periods in a mission phase can be identified and possible corrective measures evaluated. The primary purpose of the design analysis system "... is to provide data for use in comparative evaluation of alternative crew station designs". Its major values are the ease with which system changes can be evaluated. As Brown et al. state: "Any workload reduction must be evaluated in terms of the context within which this occurs and it seems senseless to increase cost by automating a feature that saves work during low workload periods only".

There are a number of other instances of the application of the synthetic methodology to the problems of workload prediction. Although the basic approaches are similar, there are some potentially important differences in detail. For example, Klein and Cassidy[38] describe an approach to estimating work requirements in which, apparently, an average required performance time is used to reflect the contribution of each task to the total work requirements, but the sum of these times can exceed the time available and thus lead to the notion of time stress. Their general procedure for analyzing the mission requirments is basically as described above. Klein and Cassidy also point out the need to recognize the nonadditivity of workload elements. This nonadditivity was investigated by evaluating a tracking task when performed in conjunction with a discrete task; they concluded: "Workload elements do not interlace in a directly additive fashion".

Wingert[39] places considerable emphasis on the fact that the performance of two tasks in combination often represents a workload that is less than the sum of the individual workloads. He used a model that took account of the nature of the task input (visual, auditory, or kinesthetic) and the task output (motor, vocal, or none required). He then prepared an "interlace table" for different combinations of two tasks with the various possible combinations of input and output modes. The actual values used in the table depended on analyses of the scanning requirements, information-processing-time predictions, and the set of summation rules assumed to apply to particular pairs of inputs and outputs. A specific set of tasks was evaluated by using a fixed-base helicopter simulator, and "interlace coefficients" were determined. The resultant coefficients are used, in the simple case, as follows:

$$\text{Total workload} = WL(1) + WL(2) - I\,WL(2)$$

where I = the interlace coefficient.

Wingert discusses the concept of interlacing in the context of parallel versus serial processing of information, and, in general, the amount of interlacing expected depends on the extent to which parallel processing is possible.

This notion of interlacing can also be viewed from the simpler time-sharing frame of reference. The highly skilled operator has typically "automated" many aspects of this complex task in a way such that many of the elements require little if any information processing (channel capacity) for satisfactory execution of the required behaviors. Consider a two-dimensional tracking as represented by the instrument landing system (ILS) display. Assume that the pilot, on approaching the outer marker, observes that he is slightly (but undesirably) below glide slope. Through long experience, he is able to apply an appropriate adjustment that will bring the aircraft smoothly to the glide slope. He *does not* then sit and watch the needle slowly drop! He turns his attention to other displays (e.g., airspeed) and knows approximately when to return his attention to the ILS display. Similarly, once he has the ILS needles centered and has established a proper rate of descent, only under very adverse conditions of wind and turbulence will he have to give the ILS display his undivided attention. In other words, how often he must look at a display to insure satisfactory performance depends on the "forcing function" acting on that display and the criticality of the task in terms of permissible error rates and amplitudes (cf. Senders[13]). To consider another kind of behavior, the neophyte automobile driver must give most of his attention to the steering task of "keeping the car on the road". For the expert driver, steering is concerned with avoiding rough spots, maintaining safe separations from oncoming traffic, etc.; keeping the car on the road has been automated. And if we look far enough we may run across an oldtime telegraph operator who can send or receive a message while simultaneously telling us about the good old days.

However, we should keep in mind that, at least at the present state-of-the-art, caution is in order in assuming too much interlacing. Such skills may be highly vulnerable to stress and other such factors (cf. Chiles, Alluisi, and Adams, (Reference 20, p.151)). By way of analogy, we do not want an aircraft designed to just withstand the maximum *expected* g and gust loads.

4.5 SIMULATION METHODS

4.5.1 Fidelity

Webster[40] defines a simulator as "one that simulates, *specif*: a device in a laboratory that enables the operator to reproduce under test conditions phenomena likely to occur in actual performance". If we interpret the word "phenomena" to mean "system-operating characteristics", then the dictionary definition certainly states the intent of the designer of the simulator. Chapanis[41] considers a simulation to be a kind of model and prefers to define models as simply being *analogies* of some particular part of the real world that is of interest to the model maker. Chapanis makes a good case for this usage, and an important value in thinking of a simulation as being an analogy is that we are all aware that analogies tend to come apart when they are pushed too hard or are examined too closely. When we talk about fidelity of simulation, we are thus talking about "how hard we can push" before the analogy breaks down.

The difficulties encountered in achieving adequate fidelity in a simulator are primarily a function of the purpose for which the simulator is to be used. Thus, for some purposes, a control stick and a display with an appropriate interface provide adequate levels of fidelity. As Hopkins[42] has said, the kinds of things that are needed on a simulator depend on "(1) your purpose in using it, and (2) your method of using it. . . . Cost effectiveness has not been demonstrated for all the bells and whistles that come as standard trimmings on our current flight training simulators".

4.5.2 Assumptions

The basic assumption underlying the use of simulation in virtually any context is that the device represents to a satisfactory degree those elements of the system being simulated that are important and relevent to the purposes of the enterprise being undertaken. More specifically, in using a simulator to study pilot workload, it is assumed that:

(1) Those factors in the real system that are relevant and important to the operator functions being evaluated are present.

(2) Those aspects of the simulation that differ from the real system will not introduce important disturbances in the measures being taken.

(3) Behavioral effects of task manipulations can be isolated from simulator operating characteristics as sources of variance.

(4) The performance effects of the variables being manipulated in the simulation do not importantly differ from the effects that would occur in the real system.

Most of the work that has focused on the evaluation of the usefulness of simulators has been done in the context of the substitution of simulator training or experience for actual flight training or experience, and even in this area many questions regarding training simulators have been at best only partially answered. (A special issue of *Human Factors* (1963, No.6) was devoted to this problem area.)

Unfortunately, many of the investigations that have looked at workload and other design questions using simulation have been reported in private company or laboratory internal publications or not at all. Thus, the open literature is virtually devoid of well-documented studies in which simulation – in the ordinary meaning of that term – was used to investigate workload; e.g., where measures were taken from the simulator to provide indices of the performance effects of workload variations as produced by changes in the simulator tasks.

4.5.3 A Flight Simulator Example

Corkindale[43] reported a study of missile control performance as a function of concurrent workload using a fixed-base flight simulator. The study included the following workload conditions:

(1) Missile control tasks only. (Two-dimensional tracking using a joy stick with the left hand and a TV display.)

(2) Simulator manual control using a Head-Up Display (HUD). (Two-dimensional tracking with control column.)

(3) Missile control plus HUD manual control. (Two, independent, two-dimensional tracking tasks – one with left hand and one with right hand. At the end of first 90 seconds, the TV came on and the subject watched for appearance of target.)

(4) Missile control task plus HUD monitoring. (Two-dimensional tracking of missile plus monitoring of HUD for an infrequently presented signal that subject responded to by pressing a button on the control column.)

Performance of the missile and aircraft control tasks was measured by recording integrated errors in each axis for each tracking task. In addition, detection time for the TV target was measured. Once the TV target was acknowledged and the crosshairs had appeared, the missile tracking task lasted just 10 seconds; the HUD aircraft control task, when present, last for approximately 3 minutes 10 seconds; the missile control task always fell in the second half of the test trial.

All but one of the measures evaluated were significantly affected by workload; surprisingly, horizontal error in tracking the TV display target was not sensitive to these workload variations. A major conclusion drawn by Corkindale[43] was that his findings fit well with the work that Rolfe[14] reviewed and interpreted to indicate that secondary tasks typically produce degradation of the performance of the primary task in spite of instructions to maintain the highest level of performance on that task. It would be interesting to know what sort of prediction the analytic method of estimating workload (e.g., Senders[13]) would make as regards the task combinations used. Corkindale cites evidence that the subjects spent a significantly smaller percentage of the time looking at the HUD when the TV was on (29.3 percent) than when the TV was off (60.3 percent), even though the HUD was the primary source of feedback to the subject as to how well he was controlling the aircraft. Therefore, one would be tempted to speculate that the analytic method would predict that a pilot cannot do both of the tasks without at least some degradation of performance on both. What, then, should we expect the pilot to do when we ask him to try to do both tasks simultaneously? Assuming that the pilots used in such a study were mission oriented, then their approach to the situation might very well be as follows:

> "This is an exercise in which I am expected to hit a target with an air-to-surface guided weapon. I have to control the missile *and* fly the airplane. I know that I cannot fly as well while controlling the missile as I can while I am not. So, I will try my best to hit the target and will consider the mission a success if I score a hit and do not crash."

It could be argued that many military pilots would follow this line of reasoning *unless* they were told that they *must maintain undiminished control* of the aircraft *even if they never hit any targets*. And with instructions of that sort, it might be difficult to maintain good levels of subject motivation to perform the task.

Assuming that Corkindale's subjects were able to handle the aircraft control task in a manner that *satisfied them* when that was their only task, what does a (significant) doubling of the error scores with the addition of the TV task mean? Did the pilots *think* they were controlling the aircraft in an acceptable manner in the two-task condition? Whether they did or not, what was their criterion? Did any of them ever "crash"? Without some sort of *absolute* error criterion, the interpretation of the results in this kind of study (or any simulator study) is very difficult. We are on somewhat firmer ground if the purpose of a study is to compare the workload properties of, for example, two alternative ways of displaying the same information. If there is a *substantial* and statistically significant advantage of one alternative, then cost-versus-effectiveness analyses can be made. But even in this simpler case, the absence of absolute criteria creates problems; for example, what procedure can be used to establish what a "substantial advantage" is in relation to "real world" requirements? In other words, we must not forget that in many important respects a simulation *is* merely an analog of some aspect of the real world.

4.5.4 A Space Simulator Example

Cotterman and Wood[44] attempted a direct treatment of the problem of criteria in a simulation context in a study of the retention of pilot skills associated with a lunar landing mission. This study involved a full mission simulation at the Martin-Marietta Corporation as a part of the NASA space program. The subjects in this study were 12 aerospace research pilots who had participated previously in a Human Reliability Program study conducted with this simulation system. The specific goal of the study reported by Cotterman and Wood was the evaluation of the retention of skill after relatively long periods (13 weeks) of disuse. The total study concerned nine separate mission phases, with from one to four performance criteria for each phase. For present purposes, only one phase will be discussed; *viz*, the "Brake and Hover" phase involved in the lunar landing.

Based on engineering analyses, permissible error rates had been established for four motion parameters during the Brake and Hover phase. These were: displacement (or range error), 200 feet; displacement rate, 10 feet/second; impact rate, 10 feet/second; percentage fuel consumed, 95 percent. Exceeding these values by appreciable amounts would incur unacceptable risk of mission failure.

The analytical approach applied by Cotterman and Wood was to use the data on the last four training trials for each pilot to establish a mean and a standard deviation for each parameter. Since their interest was in establishing whether subjects could attain performance at a high level of consistency, they selected a statistical criterion that was associated with a probability of 0.950 that the subject would perform within the criterion tolerances. The actual calculations, though somewhat laborious if done by hand, are conceptually simple. First, the standard deviation for the data from a given pilot for a given measure is computed; then, a normal deviate ("z" score) is found by dividing the difference between the criterion and the obtained score of interest by the standard deviation. A table of normal deviates can then be used to establish an approximation of the probability that the pilot in question will in fact be expected to stay within the criterion, or, using the appropriate equations, an exact probability can be computed. For one subject in the study reported by Cotterman and Wood, it was found that probabilities of staying within the criterion on the four previously mentioned variables were: 0.998; 0.525; 0.9995; and 0.9995. If the events on which each of these probabilities is based are independent, then their cumulative product is the probability that the entire mission phase will be within the criterion limit. With this approach, whether applied to simulation or to an in-flight situation – assuming that the criteria can be specified – the probabilities can be developed in a way that makes them useable for purposes of reliability engineering. The requirements are (1) the data must be quantitative in form, (2) enough repetitions per subject must be provided to achieve reasonably reliable estimates of the standard deviation, and (3) some criterion must be available that is specifiable in quantitative form.

4.6 IN-FLIGHT METHODS

4.6.1 System-Based Measures

Various techniques have been used to record indices of performance in aircraft. They have involved varying degrees of difficulty of installation and have been used with varying degrees of success. Some of the earliest systems used voltage analogs, either from direct instrument pickoffs or from repeater instruments, to drive the pens of an ink-writing oscillograph. More recently, frequency modulation techniques have been used to record analog signals onto magnetic tape; off-line computer readout and analysis can then be applied to the tapes. And still more recently, on-board digitizing techniques have been used to record data on magnetic tape directly in digital format for later computer analysis.

Some of the earliest work on studies of aircrew workload involved variations on the standard techniques of time-and-motion study (e.g., Christensen[45]), and, at about that same time, pilot workload (instrument scanning) was studied by use of motion pictures of pilot eye-movements during instrument approaches (Milton, Jones, and Fitts)[46]. Still more recently, Weir and Klein[34] describe the use of an Eye-Point of Regard system that uses a horizontal movement detector (bite board) and corneal reflection to give a resolution of "about $\pm 1°$" in either axis with respect to the eye fixation point. Photographic and videotape techniques have also been used to record general pilot activities in simulators as well as aircraft; e.g., time/frequency measures of control usage. And still more recently, Geiselhart, Shiffler, and Ivey[47] used time-and-motion study techniques in evaluating crew requirements for the KC 135 tanker aircraft on actual missions.

Roscoe and Williges[48] reported a study carried out in a Beechcraft C45H using each of eight experimental display conditions under simulated instrument flight conditions. The tasks confronting the subjects, who were naive to flying, were (1) tracking a randomly generated command flight path; (2) a disturbed attitude task that required subjects to compensate for Gaussian noise summed with the actual bank attitude signal; and (3) recovery from unusual attitudes entered with subliminal angular accelerations. All data were recorded on a strip recorder and on magnetic tape. Among other results reported by Roscoe and Williges was the finding that the maintenance of command heading was significantly better with the displays in a pursuit mode as compared to a compensatory mode.

Knoop[49] reports a study designed to evaluate the feasibility of automatically assessing T-37 student pilot performance in the Air Force Undergraduate Pilot Training program. A T-37B aircraft was instrumented to record 21 flight and control parameters in digital form on magnetic tape. Major variables (airspeed, pitch, roll, stick position in two dimensions, and rudder position) were sampled 100 times per second. Other variables, such as altitude, heading, flap position, etc., were sampled at a 10-Hz rate. A major part of this effort involved attempts on the part of instructor pilots to fly prescribed maneuvers in as nearly perfect a manner as possible. These maneuvers were broken down into phases and subjected to computer analyses in an attempt to develop measures that best characterized a high level of performance; concurrently, subjective ratings of the instructor pilots were also used as part of the evaluation. The resultant functions of the various control and performance parameters were compared with those of student pilots to try to identify those measures that best discriminate between trainees and skilled pilots. Overall, this effort met with mixed success, and major attention was diverted to trying to follow the progress of students through the training program. A major difficulty encountered was the clear lack of agreement across instructors as to what was most important in characterizing good performance in particular maneuvers.

Hasbrook, Rasmussen, and Willis[50] reported an in-flight evaluation of a "peripheral vision flight display" (PVFD) in a Beechcraft Bonanza 35A aircraft. Each of 20 pilots flew two ILS approaches with a conventional display system; they also flew five approaches with the PVFD system, but only the last two of these approaches were considered for data analysis purposes. Performance levels were recorded on a 14-channel FM analog tape system installed in the left rear seat of the aircraft. Twelve channels of information were recorded: pilot heart rate; aircraft pitch and roll (taken from the primary attitude indicator); vertical and lateral deviations from the ILS centreline (taken from the glide slope and localizer signals); altitude, airspeed, and vertical speed (obtained from the aircraft's pressure and static air systems); vertical acceleration (taken from an accelerometer located near the center of gravity of the aircraft); heading deviations (taken from a remote gyro-stabilized compass); and control wheel data (derived from mechanoelectric transducers connected to the aircraft's control cables). Event signals were inserted on a separate data channel by the use of a manual switch. Data were recorded starting at the beginning of the approach at the outer marker and ending when the runway threshold was crossed at an altitude of 100 feet; at that point the subject was instructed to increase power and go around. No differences were found between the displays, but the more experienced pilots of the group (an average of 1,267 hours of instrument time) maintained a small, significant superiority on holding to the glide slope between the outer and middle markers as compared to a less experienced group (an average of 104 hours of instrument time). Thus, although Hasbrook et al. stated that the pilots generally rated the PVFD as good to excellent, the PVFD display configuration did not result in statistically superior performance.

Billings, Gerke, and Wick[51] did a study that, though it did not involve manipulation of workload, is of interest because it involved both in-flight and simulator performance. The variable of interest was the dosage of sodium secobarbital (0, 100, or 200 mg.). The in-flight portion of the study was carried out by using a specially instrumented Cessna 172; the simulator part of the study used a GAT-1 simulator. For both the aircraft and the simulator, data were recorded in digital format at a sampling rate of 25 Hz to yield measures of average absolute error in holding to the localizer, glide path, and commanded airspeed (100 mph); root-mean-square (RMS) error was derived by appropriate computational procedures for each of the variables. The five "highly experienced professional pilots" who served in the study showed a small, nonsignificant overall increase in error across the six aircraft flights (averaged over drug conditions) and a slightly larger, *significant decrease* in error over the six simulator flights (again averaged over drug effects). It is interesting to note that whereas all of the six statistical tests carried out on the simulator data showed a significant drug effect, only four of the six tests on the aircraft data showed the drug effect to be significant. In addition, for all segments of the approach the no-drug (placebo) condition was best in the simulator, and for all but one segment the 100-mg does resulted in better performance than did the 200-mg dose in the simulator. The analogous results were mixed in the case of the aircraft data. On all three measures (glide slope, localizer, and airspeed) the RMS variability was less in the simulator than in the aircraft; and for only one absolute measure (deviation from command airspeed at the 200-mg dose) was performance in the simulator numerically poorer than in the aircraft. Direct statistical comparisons between simulator and aircraft were not reported; perhaps they were not feasible.

4.6.2 Externally Based Measures

Brictson, Ciavarelli, and Wulfeck[52] describe a system that has been used to assess the quality of aircraft carrier approaches and landings. The workload variations were those associated with night versus day landings. The procedure for recording the final approach performance involved a shipboard instrumentation system consisting of twin precision radars and a signal data recorder that provided up to eight channels of continuous flight information. The range error was reported to be on the order of 4 feet and the angular error, on the order of 0.3 milliradian. Range, true altitude, altitude error, lateral error, sink speed, true air speed, deck pitch, and closing speed were the variables usually recorded. Among other findings, Brictson et al. reported that altitude errors were greater at night than during the day with a greater tendency for the approach to be below glide slope at night. They also report that a reasonably good measure of the quality of the approach and landing was obtained by simply noting which of the four arresting wires was hooked and the number of "bolters" (no arresting wire engaged). The major difference in the tasks of night versus day landing was in the impoverishment of the visual field in terms of details of the carrier and the texture of the water. Not having those cues made the task more difficult, and Bricston et al. were able to develop differential criteria for predicting successful landings at night versus during the day for various departures from the optimum approach configuration.

4.7 DISCUSSION, RECOMMENDATIONS, CAUTIONS, AND CONCLUSIONS

4.7.1 A Hypothetical Research Vehicle

Let us assume that there exists a real aircraft system with the following capabilities: (1) An exact assignment of the nature and number of pilot duties or activities can be made for any given mission phase. (2) It is possible to vary those duties singly or in combination over time. (3) Control and display characteristics can be manipulated at will. (4) Precise and reliable quantitative indices of the task demands placed on the pilot by the system are available for all task demands placed on the pilot by the system are available for all task elements. (5) Precise and reliable quantitative measures of the skill with which the pilot meets those demands are available. (6) An adequate criterion measure of system performance is available.

What kinds of information might we expect to be able to develop as regards pilot workload through use of such a system? First, as we add tasks in different combinations, we should be able to determine the *priorities* the pilot assigns to the different tasks and whether these priority assignments are consistent across pilots; as the number of actions required per unit of time approaches and exceeds the time available, or as simultaneous demands for action arise, some tasks *will be given less attention* with a resultant lowering of performance on those tasks. Second, we should be able to determine how the different elements of the pilot's job interact as different tasks are added to the total workload; do some tasks tend to interfere with the performance of other tasks? And, third, we should be able to determine what kinds of tasks or performance functions are most sensitive to variations in total demand.

In a similar manner, for a given task load on our assumed system, we should be able to determine the relative sensitivity of the different performance demands to various environmental and procedural factors. We should, in this somewhat different context, again see which tasks are given priority. And we should be able to acquire information on the relative importance of "operator style" in system performance.

From systematic studies of task characteristics, task combinations, and procedural factors, we should be able to develop a quantitative concept of workload capacity or – as some prefer to call it – channel capacity. Thus, we should be able to arrive at a notion of workload for a given mission phase as involving some portion of the pilot's total moment-to-moment capacity to satisfy the system demands.

Unfortunately, there appear to be no instances in which a system or a simulated system has been subjected to these sorts of manipulations in any kind of programatic attack on the nature of pilot workload. [Although something like this has been done with synthetic work tasks, the programs have not been as complete or as systematic as would be desirable, and the results are, therefore, of more relevance to environmental and procedural variables than to workload *per se* (cf. Chiles et al.[20]; Alluisi[53]).]

However, we can, perhaps, make some empirically based projections (educated guesses) as to what some of the products of such a program might be. First, we would surely find that some tasks will be given priority. Which ones will depend on training and the perceived criticality of the task to the safety of the system and to the probability of mission accomplishment. For example, ILS-type guidance information will be given very high priority during very low visibility approach conditions; and there is reason to believe that some of the instruments are, on occasion, given too low a priority after breakout with potentially disastrous results.

Another predictable result is that the elements of many combinations of tasks will be found to be nonadditive (in the simplest meaning of that term). At high levels of pilot skill at time sharing, a number of tasks can apparently be performed without evidence of decrements or cross interference. However, where tasks present conflicting demands, the lack of additivity may take on a much different character; the specific effects will largely depend on the required sampling rate for the different information sources coupled with the required "dwell times"; i.e., how long it takes the pilot to extract the necessary information. Perhaps the most important single factor in this area is the degree of freedom the pilot can exercise as to exactly when various actions must be initiated.

If the suggested program were to be carried far enough, it would probably develop that only a limited number of operator styles will emerge that will allow or insure overall satisfaction of the system demands.

And, finally, it will be only after substantial and thorough research that the quantitative methods will yield readily useable indices that relate directly to "how hard the pilot has to work" with a given system workload configuration.

The fact that these above-mentioned "educated guesses" are, for the most part, rather obvious should not be allowed to detract from the clear desirability of attempting their empirical verification. Perhaps on such a "bare bones" kind of outline a general theory of workload could be developed.

4.7.2 Choosing a Method

The first and foremost factor to keep in mind in choosing a methodology in attacking some particular workload question is the purpose or goal of the research. This is true whether we are choosing from among the kinds of methods discussed in this chapter or from among those discussed in one of the other chapters.

The primary thing to keep in mind is that the measures being taken should allow the detection of operationally important changes in the pilot's ability to satisfy system demands as a function of the workload variables being manipulated. If a given measure or pattern of measures were to reveal decrements for one configuration of system demands in relation to another configuration, the decrements should be meaningful relatable to critical operational tasks in terms of pilot reliability, system safety, and/or probability of mission success. Alternatively, (and this is much more difficult to establish) if no decrements are found for a given workload configuration, it should be clearly possible to predict that the pilot could satisfy the system demands under operational conditions. At the same time, every possible effort (within reason and the scope of available resources) should be made to design the research so that maximum generality across systems is possible. Clearly, when we choose a method and select the variables that are to be measured (the dependent variables), we are committing ourselves to a particular realm of discourse as regards system workload parameters. Thus, we must be certain that the basic problem that gave rise to the research can in fact be handled within that realm of discourse. (The importance of the selection of dependent variables has been dealt with in some detail by Chapanis[54]; Alluisi[55]; and by Chiles[56,21].)

The most pressing and the most difficult problem in assessing workload effects (whatever method is chosen) lies in the development of reliable, quantitative criteria that validly reflect system performance. We need criteria against which to evaluate the results of our research. We must be able to distinguish acceptable from unacceptable, good from acceptable, and excellent from good performance of the system. We must be able to make these distinctions quantitatively and reliably. And we must be able to disentangle pilot performance, machine performance, and pilot-machine performance. Ultimately, we want a method with which it would be possible to assign reliable variance, as appropriate, to the man, to the machine, and/or to the man-machine interface.

For some specific questions this may appear to be a deceptively approachable question. For example, if we need to determine which of two instrument landing systems makes the smaller contribution to pilot workload, we could simply secure accurate measures of the deviation of the aircraft from the glide slope and the localizer and perhaps monitor airspeed. Comparison of the values of these measures for the two displays should give us an index of their workload-inducing properties. However, it is entirely conceivable that one display would lead to smaller errors only because the pilot could, by working harder, take advantage of some peculiarity of that display in holding to the proper course; at the same time, the pilot might very well be less able to respond appropriately to some emergency condition that might arise from some other quarter. Thus, in this specific example, we would need to add a variable that would shed light on how much of the pilot's workload capacity was being used up by each display. In our hypothetical, completely flexible aircraft system, we could introduce some sort of malfunction that, conceivably, could be handled readily with the otherwise poorer display but only with considerable difficulty in the case of the "better" display. This is admittedly a highly artificial example and the intent is merely to suggest a possible way in which what might appear to be a simple measurement problem might not be so easy after all. The other intent in introducing the example is to suggest that when we draw a conclusion based on a particular set of measures, the results may imply extrapolations well beyond the circumstances under which the measurements were made. (Remember, analogies, as well as examples, should not be pushed too far.)

The measurement and analysis approach described by Cotterman and Wood[44] in their evaluation of performance in a space vehicle simulator appears to show considerable promise as a technique for converting "raw" performance measurements to probabilities of meeting criterion requirements. However, there is a gap between their application and the typical pilot workload measurement situation. Specifically, in the case of the Lunar Excursion Module, the maximum values of various parameters can be specified quite readily; for example, engineering specifications dictate that the impact velocity of the vehicle on landing cannot exceed some value without risk of damage. Such precision is less clearly identifiable in the majority of aircraft operating situations; typically, rather broad latitude is possible in the flight parameters without risk of entering unsafe conditions of flight. Thus, in some areas the application of the procedure to some aircraft mission phases might become a bit arbitrary. Perhaps for research purposes it would be necessary and profitable to set up much more stringent criteria than normal, but not too stringent; the difficulty of the criteria should be such that the typical pilot from the population of pilots to which we wish to generalize would, under normal conditions, be capable of performing satisfactorily.

Assuming that we have adequate criteria of system performance that reflect both man and man-machine contributions to system output, how do we proceed?

The first step is the identification of all of those human and machine factors that could conceivably influence the variable of interest. This list typically will be unmanageable from a research point of view, *and expert judgment, based on knowledge of human behavior and system behavior*, will have to be applied to eliminate those factors of negligible or relatively small potential impact. Having developed a (presumably manageable) list of *important* factors, we attempt to phrase (or rephrase) the question such that it becomes amenable to some (as yet unspecified) research technique. We next arrange the relevant factors into two categories; one category contains items that are in the nature of constraints or boundary conditions, and the second category contains items that are in the nature of possible independent variables; this second category will, of course, include the factor or factors that gave rise to the need for the research in the first place. Now we are ready to examine the situation in detail in order to make a decision as to what would be the best research methodology to apply to the problem. At this point the available guidelines become very ambiguous and professional judgment must play a dominant role.

First, we look at what are referred to above as the boundary conditions; these are the fixed aspects of the operational system from which the problem derives; they concern factors such as the gross weight of the vehicle, its flight range, mission characteristics, number of engines, etc. Each of these factors is evaluated in relation to the question: "Might this factor be reasonably expected to have an effect on the performance in question?" Then we examine each item on the list of possible independent variables; and again we ask the question: "Might this factor be reasonably expected to have an effect on the performance in question?" Depending on the pattern of "yeses" and "noes", we will tend to direct our attention toward one methodology or another.

If, for example, the basic problem is concerned with a perceptual question, say a visual discrimination in reading two different types of dial, and kinesthetic or gravitational cues would not be expected to play a role, then perhaps a more or less traditional laboratory study might be appropriate. (We will refer to this study as task A.) However, if the instrument reading must be made while performing some other task, say a two-dimensional tracking task (we will call this study task B), then perhaps a part-task simulator would be in order. If the performance of task B may be importantly influenced by the insertion of command information, then a more elaborate simulation might be in order (task C). And if kinesthetic cues may be important, we may need to go to a motion-type simulator or perhaps an in-flight evaluation.

Finally, we must select the dependent variable – the thing we are going to measure. This may be a time measure: how fast can the pilot do a task? It may be an absolute error measure: how often did he hit the wrong switch? It may be a relative error measure: what was his average deviation from glide slope? Whatever the measure, we should if at all possible try to relate the findings back to system-relevant criteria developed in a manner analogous to that described by Cotterman and Wood[44]. All too often, the thing that is chosen for measurement is that which is easiest to acquire or has been used most often in the past without any specific rationale having been shown that relates the measure to real-system performance questions.

In some cases the results of the study (accuracy of dial reading in the above-described example) may provide information that is more or less directly interpretable in terms of workload. But what if there is no change in any of the measures as a function of which dial is used? Can we infer that the two dials represent equal workload contributions? The answer is, of course, no. Only after we have pushed the total workload to a maximum reasonable and likely level and found no differences on any measures should we be willing to *assume* the equality of the two displays. (It is a peculiarity of statistical methodology that we cannot *prove* they are equal.) The procedure we use to push the apparent level of workload to a maximum is, again, a matter of professional judgment. *But it is an extremely important judgment.* If workload is added in an obviously artificial manner, especially if our subjects are operational personnel, we may lose them – motivationally speaking. We must always be sure that the research situation – be it laboratory, simulator, or aircraft – is presented in a manner such that it will be responded to as a "real" situation as opposed to a game or a contrived – and thus (perhaps) meaningless – exercise.

Let no one make the mistake of assuming that this process of choosing a method is easily executed. The problems are many and the decisions difficult.

4.7.3 Conclusions

The general approaches that we have labelled "laboratory methods" in this chapter are probably best suited to conducting background research on more general questions pertaining to workload. Wherever they are *appropriate* they are the method of choice because of the typically high degree of control possible and the attendant high levels of reliability. The synthetic work method is especially well suited to examining general workload questions because, by its nature, tasks can be added, removed, and modified with relative ease, and, depending on the overall level of complexity, large investments in training time are not required. The fact that it does not simulate an aircraft is both a strength and a weakness; it is a weakness because of problems of generalizing to specific systems; it is a strength because, if the tasks are well chosen, operational subjects can fairly easily be convinced to react to the synthetic work device for what it is and not make unfavorable comparisons between its behavior and the behavior of an aircraft. The secondary loading task method, especially when applied in a simulation or in-flight context, must be used with care. First, the task that is used to produce the load increments must be somehow (at least rationally) relatable to the kinds of activities it is presumed to assess in relation to the real system. Second, the properties of this task itself must be examined; at a minimum its reliability and relation to other tasks should be known. Although some authors (e.g. Rolfe[14], and Corkindale[43]) argue that the primary task should remain unaffected by the introduction of the loading task, this condition appears to be unnecessarily restrictive. If the loading task is properly selected (as noted above) and contradictory results are obtained (e.g., primary task A shows a decrement, primary task B is unchanged, but the loading task shows a decrement with B and not with A), the findings may be of little relevance to workload (or channel) capacity as a unitary concept; however, if such results were not simply the product of some uncontrolled condition, the finding would certainly be of theoretical if not practical interest. Perhaps it is better at this stage of development to consider the concepts *channel capacity* and *single channeledness* as being merely manners of speaking and serving primarily as heuristic devices. Although this does not argue against the ultimate possibility that the operator is single channelled, present evidence suggests that the information-handling capacity of the human operator is influenced by too great a variety of factors to try to permanently settle the single-channel hypothesis at this time. Returning to and slightly changing the above example, if task A shows more decrement than task B with the addition of the loading task and the loading task is performed better with task B than with task A, we certainly have learned something about the workload properties of the tasks. The findings, of course, remain ambiguous as regards channel capacity.

The analytic and the synthetic methods both appear to yield reasonable results, but both techniques rest on relatively fragile data bases. With further research on what may be called time sharing behavior, or what Wingert[39] calls function interlacing, the synthetic method promises to be a very useful aid in the design of systems and the allocation of workload. There is, however, considerable risk that the detailed task information required to apply the method will be collected and stored in a manner that will tend to limit its distribution and result in substantial amounts of unnecessary duplication of effort. Previous attempts to develop clearing houses for the information have not met with noteworthy success.

Simulators, especially those controlled by general purpose digital computers, have the *potential* of generating large amounts of very useful information on workload. However, whether the programs that resulted in their acquisition will allow adequate access to such systems for research purposes remains to be seen. But even given adequate access, research with simulators is not without its problems. First, naive subjects cannot be expected to learn to fly in a matter of a few hours; therefore, for most purposes – or at least for those purposes in which the full capability of the simulator is used – trained pilots are required who have adequate experience with that simulator and/or the aircraft it simulates. Thus, salaries can become a significant part of any substantial research effort. Second, the simulator is, first and foremost designed and built to *appear* to behave like the aircraft it simulates; the quality of the signals internal to the simulator need not be very high to satisfy that requirement. Thus, especially with the older simulators, the available signals often introduce an unacceptably high degree of unreliability in the final measures. Third, because the simulator is designed to mimic the airplane, many of the functions are interconnected in such a way that it can be very difficult to separate them out. For example, the relative contributions of the simulator, present performance of interest, concurrent performance that is not of direct interest and the interactions of these factors as sources of variance may be hopelessly entangled. And, fourth, also because the simulator is designed to mimic a particular airplane, generalization to other aircraft with significantly different characteristics (such as panel layout and operating procedures) becomes rather difficult.

Except for some of the safety limitations, in-flight methods can be used on virtually any problem suitable for investigation in a simulator. However, the recording of data of demonstrated reliability is a significant problem. Generally speaking, aircraft are electrically very noisy, and, where magnetic tape recordings are made (either digitally or through frequency modulation techniques), substantial programing for signal "reconditioning" is typically required; *glitches* are a constant source of annoyance[49]. Unfortunately, no reports of reliability data have been discovered for in-flight recorded performance measures or for simulator performance measures. In fact, this is a major technical deficiency in virtually all the reported research using these two methods. (This criticism applies equally well to much of the other reported research related to the measurement of workload; *viz*, laboratory research.)

Some readers may be disappointed that firmer guidelines have not been offered as to how to design and conduct research on workload problems in aviation operations. Those who are familiar with the behavioral literature on the measurement of complex human performance will understand the absence of precise, "cookbook" rules for proceeding.

REFERENCES

1. Gerthewohl, S.J. — *Definition and Measurement of Perceptual and Mental Workload in Aircrews and Operators of Air Force Weapon Systems: A Status Report*. AGARD Conf. Preprint No.181, Higher Mental Functioning in Operational Environments.

2. Gartner, W.B.; Murphy, M.R. — *Pilot Workload and Fatigue: A Critical Survey of Concepts and Assessment Techniques*. National Aeronautics and Space Administration Report No. NASA-TN-D-8365, November 1976.

3. White, R.T. — *Task Analysis Methods: Review and Development of Techniques for Analyzing Mental Workload in Multiple-Task Situations*. McDonnel Douglas Corporation, Report No. MDC J5291, September 1971.

4. Jahns, D.W. — *Operator Workload: What Is It and How Should It Be Measured*? In K.D.Gross and J.J.McGrath (Eds.), Crew System Design. Santa Barbara, California, Anacapa Sciences, Inc., pp.281-288, July 1973.

5. Spyker, D.A.; Stackhouse, S.P.; Khalafalla, A.S.; McLane, R.C. — *Development of Techniques for Measuring Pilot Workload*. National Aeronautics and Space Administration Report No. NASA-CR-1888, November 1971.

6. Bartlett, F. — *The Measurement of Human Skill*. British Medical Journal, 4510: pp.835-838 and 4511: pp.877-880, 1947.

7. Conrad, R. — *Adaptation to Time in a Sensorimotor Skill*. J. Experimental Psychology, 49: pp.115-121, 1955.

8. Conrad, R. — *The Timing of Signals in Skill*. J. Experimental Psychology, 51: pp.365-370, 1956.

9. Knowles, W.B.; Garvey, W.D.; Newlin, E.P. — *The Effect of Speed and Load on Display-Control Relationships*. J. Experimental Psychology, 46: pp.65-75, 1953.

10. Knowles, W.B. — *Operator Loading Tasks*. Human Factors, 5: pp.155-161, 1963.

11. Knowles, W.B.; Rose, D.J. — *Manned Lunar Landing Simulation*. Paper presented at IEEE National Winter Convention on Military Electronics, Los Angeles, California, 1963.

12. Benson, A.J.; Huddleston, H.F.; Rolfe, J.M. — *A Psychophysiological Study of Compensatory Tracking on a Digital Display*. Human Factors, 7: pp.457-472, 1965.

13. Senders, J.W. — *The Estimation of Operator Workload in Complex Systems.* In K.B.DeGreen (Ed.), Systems Psychology, pp.207-217. New York, McGraw Hill, 1970.

14. Rolfe, J.M. — *The Secondary Task as a Measure of Mental Load.* In W.T.Singleton, J.G.Fox, and D.Whitfield (Eds.): Measurement of Man at Work, pp.135-148. Taylor and Francis Ltd., London, 1971.

15. Kelley, C.R.; Wargo, M.J. — *Cross-Adaptive Operator Loading Tasks.* Human Factors, 9: pp.395-404, 1967.

16. O'Donnell, R.D. — *Secondary Task Assessment of Cognitive Workload in Alternative Cockpit Configurations.* AGARD Conf. Preprint No.181, Higher Mental Functioning in Operational Environments. London, 1975.

17. Sternberg, S. — *High Speed Scanning in Human Memory.* Science, 153: pp.652-654, 1966.

18. Sternberg, S. — *Memory Scanning: New Findings and Current Controversies.* Quarterly Journal of Experimental Psychology, 27: pp.1-32, 1975.

19. Darley, C.F.; Klatzky, R.L.; Atkinson, R.C. — *Effects of Memory Load on Reaction Time.* J. Experimental Psychology, 96: pp.232-234, 1972.

20. Chiles, W.D.; Alluisi, E.A.; Adams, O.S. — *Work Schedules and Performance During Confinement.* Human Factors, 10: pp.143-196, 1968.

21. Chiles, W.D. — *Complex Performance: The Development of Research Criteria Applicable in the Real World.* In W.T.Singleton, J.G.Fox, and D.Whitfield (Eds.): Measurement of Man at Work, pp.159-164. Taylor and Francis, Ltd., London, 1971.

22. Hall, T.J.; Passey, G.E.; Meighan, T.W. — *Performance of Vigilance and Monitoring Tasks as a Function of Workload.* Aerospace Medical Research Laboratories Report No. AMRL-TR- 65-22, Wright-Patterson AFB, Ohio, 1965

23. Jennings, A.E.; Chiles, W.D. — *An Investigation of Time-Sharing Ability as a Factor in Complex Performance.* Human Factors (in press).

24. Chiles, W.D.; Jennings, A.E. — *Effects of Alcohol on Complex Performance.* Human Factors, 12: pp.605-612, 1970.

25. Senders, J.W. — *Man's Capacity to Use Information from Complex Displays.* In H.Quastler (Ed.), Information Theory in Psychology. The Free Press, Glencoe, Illinois, 1955.

26. Senders, J.W. — *Tracking with Intermittent Stimuli (Forced Sampling).* Air Research and Development Report No.ARDC-TR-56-8,1956.

27. Senders, J.W. — *Information Input Rates to Human Users: Recent Research Results.* WADC Symposium on Air Force Flight Instrumentation Program. Wright-Patterson AFB, Ohio, 1958.

28. Senders, J.W. — *The Human Operator as a Monitor and Controller of Multi-degree of Freedom Systems.* IEEE Transactions on Human Factors in Electronics, Report No.HFE-5(1), pp.2-5, 1964.

29. Senders, J.W.; Stevens, K.A. — *Re-analysis of the Pilot Eye-Movement Data.* Bolt, Beranek, and Newman, Inc., Report No.1136, Cambridge, Massachusetts, 1964.

30. Carbonell, J.R.; Ward, J.L.; Senders, J.W. — *A Queueing Model of Visual Sampling Experimental Validation.* IEEE Transaction on Man-Machine Systems, MMS-9, No.3, pp.82-87, 1968.

31. Clement, W.F.; Jex, H.R.; Graham, D. — *A Manual Control-Display Theory Applied to Instrument Landing of a Jet Transport.* IEEE Transactions on Man-Machine Systems, MMS-9, pp.93-110, 1968.

32. McRuer, D.T.; Jex, H.R. — *A Systems Analysis Theory of Manual Control Displays.* Proc. 3rd Annual NASA-University Conf. on Manual Control, Report No. NASA-SP-144, pp.9-28, 1967.

33. McRuer, D.T. Jex, H.R. Clement, W.F. Graham, D. — *A Systems Analysis Theory for Display in Manual Control.* Systems Technology, Inc., Technical Report No. TR-163-1, June 1968.

34. Weir, D.H. Klein, R.H. — *The Measurement and Analysis of Pilot Scanning and Control Behavior During Simulated Instrument Approaches.* National Aeronautics and Space Administration Report No. NASA-CR-1535, 1970.

35. Allen, R.W. Clement, W.F. Jex, H.F. — *Research on Display Scanning, Sampling, and Reconstruction Using Separate Main and Secondary Tasks.* National Aeronautics and Space Administration Report No. NASA-CR-1569, 1970.

36. Munger, S.J. Smith, R.W. Payne, D. — *An Index of Electronic Equipment Operability – Data Store.* American Institutes for Research, Pittsburg, Pennsylvania, 1967.

37. Brown, E.L. Stone, G. Pearce, W.E. — *Improving Cockpits Through Crew Workload Measurement.* Douglas AC Corporation Report No. MDC 63-55, 1975.

38. Klein, T.J. Cassidy, W.B. — *Relating Operator Capabilities to System Demands.* Proceedings of the 16th Annual Meeting of the Human Factors Society, pp.324-334, October 1972.

39. Wingert, J.W. — *Function Interlace Modifications to Analytic Workload Predictions.* In K.D.Gross and J.D.McGrath (Eds.), Crew System Design. Anacapa Sciences, Inc., Santa Barbara, California, 1973.

40. – — *Webster's Third New International Dictionary of the English Language, Unabridged,* P.B.Gove (Ed.), G & C Merriam Company, Springfield, Massachusetts, 1967.

41. Chapanis, A. — *Men, Machines, and Models.* American Psychologist, 16: pp.113-131, 1961.

42. Hopkins, C.O. — *How Much Should You Pay For That Box?* Human Factors, 17; pp.533-541, 1975.

43. Corkindale, K.G.G. — *A Flight Simulator Study of Missile Control Performance as a Function of Concurrent Workload.* In AGARD Conf. Proc. No.146, *Simulation and Study of High Workload Operations,* pp.A5-1 to A5-6, 1974.

44. Cotterman, T.E. Wood, M.E. — *Retention of Simulated Lunar Landing Skills: A Test of Pilot Reliability.* Aerospace Medical Research Laboratories Report No. AMRL-TR-66-222, Wright-Patterson AFB, Ohio, 1967.

45. Christensen, J.M. — *Aerial Analysis of Navigator Duties with Special Reference to Equipment Design and Workspace Layout. II. Navigator and Radar Operator Duties During Three Arctic Missions.* USAF Report No. MCREXD-694-15A, February 1948.

46. Milton, J.L. Jones, R.E. Fitts, P.M. — *Eye Fixations of Aircraft Pilots: II. Frequency, Duration and Sequence of Fixations when Flying the USAF Instrument Low Approach System (ILAS).* Report No. USAF-TR-5839, Wright-Patterson AFB, Ohio, October 1949.

47. Geiselhart, R. Schiffler, R.J. Ivey, L.J. — *A Study of Task Loading Using a Three Man Crew on a KC-135 Aircraft.* Aeronautical Systems Division Report No. ASD-TR-76-19, Wright-Patterson AFB, Ohio, October 1976.

48. Roscoe, S.N. Williges, R.C. — *Motion Relationships in Aircraft Attitude and Guidance Displays: A Flight Experiment.* Human Factors, 17: pp.374-387, 1975.

49. Knoop, P.A. — *Advanced Instructional Provisions and Automated Performance Measurement.* Human Factors, 15: pp.583-597, 1973.

50. Hasbrook, A.H. Rasmussen, P.G. Willis, D.M. — *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, November 1975.

51. Billings, C.E.; Gerke, R.J.; Wick, R.L. — *Comparisons of Pilot Performance During Simulated and Actual Flight*. Aviation, Space, and Environmental Medicine, 46: pp.304-308, 1975.

52. Brictson, C.A.; Ciavarelli, A.P.; Wulfeck, J.W. — *Operational Measures of Aircraft Carrier Landing System Performance*. Human Factors, 11: pp.281-289, 1969.

53. Alluisi, E.A. — *Methodology in the Use of Synthetic Tasks to Assess Complex Performance*. Human Factors, 9: pp.375-384, 1967.

54. Chapanis, A. — *The Search for Relevance in Applied Research*. In W.T.Singleton, J.G.Fox, and D.Whitfield (Eds.): Measurement of Man at Work, pp.1-14. Taylor and Francis Ltd., London, 1971.

55. Alluisi, E.A. — *Optimum Uses of Psychobiological, Sensorimotor, and Performance Measurement Strategies*. Human Factors, 17: pp.309-320, 1975.

56. Chiles, W.D. — *Assessment of the Performance Effects of the Stresses of Space Flight*. Aerospace Medical Research Laboratories Report No. AMRL-TR-66-192, Wright-Patterson AFB, Ohio, December 1966.

BIBLIOGRAPHY

Barber, M.R.; Jones, C.K.; Sisk, T.R.; Haise, F.W. — *An Evaluation of the Handling Qualities of Seven General-Aviation Aircraft*. National Aeronautics and Space Administration Report No. NASA-TN-D-3726, November 1966.

Belcher, J.J. — *A Technique for Assessing Operability/Effectiveness of Control-Display Systems*. In K.D.Grass and J.J.McGrath (Eds.), Crew System Design, pp.241-249. Anacapa Sciences, Inc., Santa Barbara, California, July 1973.

Bergström, B. — *Complex Psychomotor Performance During Different Levels of Experimentally Induced Stress Pilots*. In L.Levi (Ed.), Emotional Stress, pp.239-250. American Elsevier Publishing Company, New York, 1967.

Billings, C.E.; Wick, R.L.; Gerke, R.J.; Chase, R.C. — *The Effects of Alcohol on Pilot Performance During Instrument Flight*. FAA Office of Aviation Medicine Report No. FAA-AM-72-4, January 1972.

Brictson, C.A. — *Operational Measures of Pilot Performance During Final Approach to Carrier Landing*. In AGARD Conf. Proc. No.56, Measurement of Aircrew Performance, the Flight Deck Workload and Its Relation to Pilot Performance, pp.7-1 to 7-11, December 1969.

Chapanis, A; Van Cott, H.P. — *Human Engineering Tests and Evaluations*. In H.P. Van Cott and R.G.Kinkade (Eds.), Human Engineering Guide to Equipment Design, pp.701-728. McGraw-Hill Company, Washington, D.C., 1972.

Connelly, E.M.; Bourne, F.J.; Loental, D.G.; Migliaccio, J.S.; Burchick, D.A.; Knoop, P.A. — *Candidate T-37 Pilot Performance Measures for Five Contact Manoeuvres*. Brooks AFB, Texas; Air Force Systems Command Report No. AFHRL-TR-74-88, December 1974.

Fraser, T.M. — *Philosophy of Simulation in a Man-Machine Space Mission System*. National Aeronautics and Space Administration Report No. NASA-SP-102, 1966.

Frost, G. — *Man-Machine Dynamics*. In H.P. Van Cott and R.G.Kinkade (Eds.): Human Engineering Guide to Equipment Design, pp.227-310. McGraw-Hill Company, Washington, D.C., 1972.

Gagne, R.M.
Melton, A.W.
Psychological Principles in System Development. Holt, Rinehart, and Winston, New York, 1962.

Geiselhart, R.
Koeteeuw, R.I.
Schiffler, R.J.
A Study of Task Loading Using a Four-Man Crew on a KC-135 Aircraft (Giant Boom). Aeronautical Systems Division Report No. ASD-TR-76-33, Wright-Patterson AFB, Ohio, April 1977.

Hall, E.R.
Parker, J.F., Jr
Meyer, D.E.
A Study of Air Force Flight Simulator Programs. Aerospace Medical Research Laboratories Report No. AMRL-TR-67-111, Wright-Patterson AFB, Ohio, June 1967.

Hasbrook, A.H.
Rasmussen, P.G.
Aural Glide Slope Cues: Their Effect on Pilot Performance During In-Flight Simulated ILS Instrument Approaches. FAA Office of Aviation Medicine Report No. FAA-AM-71-24, May 1971.

Hasbrook, A.H.
Rasmussen, P.G.
In-flight Performance of Civilian Pilots Using Moving-Aircraft and Moving-Horizon Attitude Indicators. FAA Office of Aviation Medicine Report No. FAA-AM-73-9, June 1973.

Henry, P.H.
Flueck, J.A.
Sanford, J.F.
Keiser, H.N.
McNee, R.C.
Walter, W.H., III
Webster, K.H.
Hartman, B.O.
Lancaster, M.C.
Assessment of Performance in a Link GAT-1 Flight Simulator at Three Alcohol Dose Levels. Aerospace Medicine, 45: pp.33-44, 1974.

Hodge, D.C. (Ed.)
Standardization of Tasks and Measures for Human Factors Research (Proceedings of a conference held at Texas Tech University). Aberdeen Research and Development Center Report No. TM-19-17, Aberdeen Proving Ground, Maryland, March 1970.

Howell, W.C.
Goldstein, I.L.
Engineering Psychology, Current Perspectives in Research. New York, Meredith Corporation, 1971.

Keenan, J.J.
Parker, T.C.
Lenzycki, H.P.
Concepts and Practices in the Assessment of Human Performance in Air Force Systems. Aerospace Medical Research Laboratories Report No. AMRL-TR-65-168, Wright-Patterson AFB, Ohio, September 1965.

Kidd, J.S.
A New Look at System Research and Analysis. Human Factors, 4: pp.209-216, 1962.

Kitchin, J.B.
Graham, A.
Mental Loading of Process Operators: An Attempt to Devise a Method of Analysis and Assessment. Ergonomics, 4: pp.1-15, 1961.

Knowles, W.B.
Aerospace Simulation and Human Performance Research. Human Factors, 9; pp.149-159, 1967.

Kraft, C.L.
Elworth, C.L.
Flight Deck Workload and Night Visual Approach Performance. In AGARD Conf. Proc. No.56, Measurement of Aircrew Performance, the Flight Deck Workload and its Relation to Pilot Performance, pp.11-1 to 11-14, December 1969.

Levison, W.H.
Elkind, J.I.
Ward, J.L.
Studies of Multivariable Manual Control Systems: A Model for Task Interference. National Aeronautics and Space Administration Report No. NASA-CR-1746, May 1971.

Meister, D.
Methods of Predicting Human Reliability in Man-Machine Systems. Human Factors, 6: pp.621-646, 1964.

Milligan, J.R.
A Student Pilot Automatic Monitoring System. In K.D.Gross and J.J.McGrath (Eds.), Crew System Design, pp.301-310. Anacapa Sciences, Inc., Santa Barbara, California, July 1973.

Parsons, H.M.
Man-Machine System Experiments. The John Hopkins Press, Baltimore, 1972.

Povenmire, H.K.
Roscoe, S.N.
An Evaluation of Ground-Based Flight Trainers in Routine Primary Flight Training. Human Factors, 13: 109-116, 1971.

Rolfe, J.M., (Ed.) — *Vehicle Simulation for Training and Research*. Farnborough, Hants, RAF Institute of Aviation Medicine Report No. IAM-442, March 1968.

Roscoe, S.N. — *Assessment of Pilotage Error in Airborne Area Navigation Procedures*. Human Factors, 16: pp.223-228, 1974.

Schiffler, R.J.
Geiselhart, R.
Ivey, L. — *Crew Composition Study for an Advanced Tanker/Cargo Aircraft (ATCA)*. Deputy for Engineering, Aeronautical Systems Division Report No. ASD-TR-76-20, Wright-Patterson AFB, Ohio, October 1976.

Schohan, B.
Rawson, H.E.
Soliday, S.M. — *Pilot and Observer Performance in Simulated Low Altitude High Speed Flight*. Human Factors, 7: pp.257-265, 1965.

Schultz, W.C.
Newell, F.D.
Whitbeck, R.F. — *A Study of Relationships Between Aircraft System Performance and Pilot Ratings*. National Aeronautics and Space Administration Report No. NASA-CR-1643, July 1970.

Shackel, B. — *Man-Computer Interaction – The Contribution of the Human Sciences*. Ergonomics, 12: pp.485-499, 1969.

Shannon, R.H.
Waag, W.L. — *Toward the Development of a Criterion for Fleet Effectiveness in the F-4 Fighter* Community. Proceedings of the 16th Annual Meeting of the Human Factors Society, pp.335-340, October 1972.

Saliday, S.M.
Schohan, B. — *Task Loading of Pilots in Simulated Low-Altitude High-Speed Flight*. Human Factors, 7: pp.45-53, 1965.

Wherry, R.J., Jr — *Human Performance Studies for the Airborne Crew Station Design Process*. In K.D.Gross and J.J.McGrath (Eds.), Crew System Design, pp.21-28. Anacapa Sciences, Inc., Santa Barbara, California, July 1973.

Zaitzeff, L.P. — *Aircrew Task Loading in the Boeing Multimission Simulator*. In AGARD Conf. Proc. No.56, Measurement of Aircrew Performance, the Flight Deck Workload, and its Relation to Pilot Performance, pp.8 to 8-3, December 1969.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
(14)	AGARD-AG-233	ISBN 92-835-74-X	UNCLASSIFIED

5. Originator

Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly sur Seine, France

6. Title

(6) ASSESSING PILOT WORKLOAD,

7. Presented at

8. Author(s)	9. Date
(10) W. Dean/Chiles, George A. / Ellis Alan H./Roscoe ~~Various~~	(11) February 1978

10. Author's Address	11. Pages
Various	84 (12) 85p

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

Pilots (personnel)	Human factors engineering	Stress (psychology)
Work measurement	Design criteria	
Performance evaluation	Stress (physiology)	

14. Abstract

The assessment of levels of pilot workload associated with the various phases and sub-phases of flight is important in the design, development, and evaluation of aircraft handling qualities and of display and guidance systems. This AGARDograph, written primarily for flight test engineers and pilots, is intended as a guide to the different methods available for estimating workload and in particular to those techniques suitable for use in aircraft. An introductory chapter briefly reviews the various concepts and classifications of workload; the former tend to fall into two main areas, those related to workload as task-demands and those to workload as pilot-effort. In Chapter 2, subjective assessment, at present the most used method, is discussed from the viewpoint of the test pilot. Physiological methods in general are reviewed in Chapter 3 with those techniques available for use in flight being discussed in more detail. Chapter 4 describes various objective methods and presents examples of their practical application. Whereas the methods in Chapter 2 and 3 are appropriate only to workload as effort, objective methods contain techniques appropriate to workload as task-demands as well as to effort. The former techniques are particularly valuable for providing data which can be used to construct models and to predict levels of workload. Different modelling techniques will be discussed in a proposed supplement entitled Engineering Methods.

This AGARDograph was prepared at the request of the Flight Mechanics Panel of AGARD.

400043 TC

AGARDograph No.233
Advisory Group for Aerospace Research and Development, NATO
ASSESSING PILOT WORKLOAD
Published February 1978
84 pages

The assessment of levels of pilot workload associated with the various phases and sub-phases of flight is important in the design, development, and evaluation of aircraft handling qualities and of display and guidance systems. This AGARDograph, written primarily for flight test engineers and pilots, is intended as a guide to the different methods available for estimating workload and in particular to those techniques suitable for use in aircraft. An introductory chapter briefly reviews the

P.T.O.

AGARD-AG-233

Pilots (personnel)
Work measurement
Performance evaluation
Human factors engineering
Design criteria
Stress (physiology)
Stress (psychology)

AGARDograph No.233
Advisory Group for Aerospace Research and Development, NATO
ASSESSING PILOT WORKLOAD
Published February 1978
84 pages

The assessment of levels of pilot workload associated with the various phases and sub-phases of flight is important in the design, development, and evaluation of aircraft handling qualities and of display and guidance systems. This AGARDograph, written primarily for flight test engineers and pilots, is intended as a guide to the different methods available for estimating workload and in particular to those techniques suitable for use in aircraft. An introductory chapter briefly reviews the

P.T.O.

AGARD-AG-233

Pilots (personnel)
Work measurement
Performance evaluation
Human factors engineering
Design criteria
Stress (physiology)
Stress (psychology)

AGARDograph No.233
Advisory Group for Aerospace Research and Development, NATO
ASSESSING PILOT WORKLOAD
Published February 1978
84 pages

The assessment of levels of pilot workload associated with the various phases and sub-phases of flight is important in the design, development, and evaluation of aircraft handling qualities and of display and guidance systems. This AGARDograph, written primarily for flight test engineers and pilots, is intended as a guide to the different methods available for estimating workload and in particular to those techniques suitable for use in aircraft. An introductory chapter briefly reviews the

P.T.O.

AGARD-AG-233

Pilots (personnel)
Work measurement
Performance evaluation
Human factors engineering
Design criteria
Stress (physiology)
Stress (psychology)

AGARDograph No.233
Advisory Group for Aerospace Research and Development, NATO
ASSESSING PILOT WORKLOAD
Published February 1978
84 pages

The assessment of levels of pilot workload associated with the various phases and sub-phases of flight is important in the design, development, and evaluation of aircraft handling qualities and of display and guidance systems. This AGARDograph, written primarily for flight test engineers and pilots, is intended as a guide to the different methods available for estimating workload and in particular to those techniques suitable for use in aircraft. An introductory chapter briefly reviews the

P.T.O.

AGARD-AG-233

Pilots (personnel)
Work measurement
Performance evaluation
Human factors engineering
Design criteria
Stress (physiology)
Stress (psychology)

various concepts and classifications of workload; the former tend to fall into two main areas, those related to workload as task-demands and those to workload as pilot-effort. In Chapter 2, subjective assessment, at present the most used method, is discussed from the viewpoint of the test pilot. Physiological methods in general are reviewed in Chapter 3 with those techniques available for use in flight being discussed in more detail. Chapter 4 describes various objective methods and presents examples of their practical application. Whereas the methods in Chapter 2 and 3 are appropriate only to workload as effort, objective methods contain techniques appropriate to workload as task-demands as well as to effort. The former techniques are particularly valuable for providing data which can be used to construct models and to predict levels of work-load. Different modelling techniques will be discussed in a proposed supplement entitled Engineering Methods.

This AGARDograph was prepared at the request of the Flight Mechanics Panel of AGARD.

ISBN 92-835-74-X

various concepts and classifications of workload; the former tend to fall into two main areas, those related to workload as task-demands and those to workload as pilot-effort. In Chapter 2, subjective assessment, at present the most used method, is discussed from the viewpoint of the test pilot. Physiological methods in general are reviewed in Chapter 3 with those techniques available for use in flight being discussed in more detail. Chapter 4 describes various objective methods and presents examples of their practical application. Whereas the methods in Chapter 2 and 3 are appropriate only to workload as effort, objective methods contain techniques appropriate to workload as task-demands as well as to effort. The former techniques are particularly valuable for providing data which can be used to construct models and to predict levels of work-load. Different modelling techniques will be discussed in a proposed supplement entitled Engineering Methods.

This AGARDograph was prepared at the request of the Flight Mechanics Panel of AGARD.

ISBN 92-835-74-X

various concepts and classifications of workload; the former tend to fall into two main areas, those related to workload as task-demands and those to workload as pilot-effort. In Chapter 2, subjective assessment, at present the most used method, is discussed from the viewpoint of the test pilot. Physiological methods in general are reviewed in Chapter 3 with those techniques available for use in flight being discussed in more detail. Chapter 4 describes various objective methods and presents examples of their practical application. Whereas the methods in Chapter 2 and 3 are appropriate only to workload as effort, objective methods contain techniques appropriate to workload as task-demands as well as to effort. The former techniques are particularly valuable for providing data which can be used to construct models and to predict levels of work-load. Different modelling techniques will be discussed in a proposed supplement entitled Engineering Methods.

This AGARDograph was prepared at the request of the Flight Mechanics Panel of AGARD.

ISBN 92-835-74-X

various concepts and classifications of workload; the former tend to fall into two main areas, those related to workload as task-demands and those to workload as pilot-effort. In Chapter 2, subjective assessment, at present the most used method, is discussed from the viewpoint of the test pilot. Physiological methods in general are reviewed in Chapter 3 with those techniques available for use in flight being discussed in more detail. Chapter 4 describes various objective methods and presents examples of their practical application. Whereas the methods in Chapter 2 and 3 are appropriate only to workload as effort, objective methods contain techniques appropriate to workload as task-demands as well as to effort. The former techniques are particularly valuable for providing data which can be used to construct models and to predict levels of work-load. Different modelling techniques will be discussed in a proposed supplement entitled Engineering Methods.

This AGARDograph was prepared at the request of the Flight Mechanics Panel of AGARD.

ISBN 92-835-74-X

4

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 OZ2

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
Postfach 860880
D-8 München 86

GREECE
Hellenic Armed Forces Command
D Branch, Athens

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
Direccao do Servico de Material
da Forca Aerea
Rua de 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 22151, 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 22151, USA

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

ISBN 92-835-74-X