NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

AC/323(HFM-104)TP/48

www.rta.nato.int

RTO TECHNICAL REPORT **TR-HFM-104**

Operator Functional State Assessment

(L'évaluation de l'aptitude opérationnelle de l'opérateur humain)

This Technical Report has been prepared by the RTO Human Factors and Medicine Panel (HFM) Task Group HFM-056/TG-008.

The material in this publication also supported a Lecture Series under the sponsorship of the Human Factors and Medicine Panel (HFM) presented on 8-9 September 2003 in Kiev, Ukraine; 11-12 September 2003 in Brussels, Belgium; and 2-3 October 2003 in San Diego, USA.

Published February 2004

Distribution and Availability on Back Cover

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

AC/323(HFM-104)TP/48

www.rta.nato.int

RTO TECHNICAL REPORT **TR-HFM-104**

Operator Functional State Assessment

(L'évaluation de l'aptitude opérationnelle de l'opérateur humain)

This Technical Report has been prepared by the RTO Human Factors and Medicine Panel (HFM) Task Group HFM-056/TG-008.

The material in this publication also supported a Lecture Series under the sponsorship of the Human Factors and Medicine Panel (HFM) presented on 8-9 September 2003 in Kiev, Ukraine; 11-12 September 2003 in Brussels, Belgium; and 2-3 October 2003 in San Diego, USA.

The Research and Technology Organisation (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote co-operative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective co-ordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also co-ordinates RTO's co-operation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of co-operation.

The total spectrum of R&T activities is covered by the following 7 bodies:

- AVT Applied Vehicle Technology Panel
- HFM Human Factors and Medicine Panel
- IST Information Systems Technology Panel
- NMSG NATO Modelling and Simulation Group
- SAS Studies, Analysis and Simulation Panel
- SCI Systems Concepts and Integration Panel
- SET Sensors and Electronics Technology Panel

These bodies are made up of national representatives as well as generally recognised 'world class' scientists. They also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier co-operation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

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

Published February 2004

ISBN 92-837-1111-4

Operator Functional State Assessment

(RTO-TR-HFM-104)

Executive Summary

The human operator is a crucial component of complex modern systems. The complexity of these systems, the rapid tempo of contemporary military operations and reduced staffing all contribute to the high cognitive demands experienced by military personnel. Unfortunately, these systems do not account for the functional state of the human operator. The rate of information flow, the number of decisions, and actions that must be carried out can become so demanding that the cognitive capacity of the operator can be exceeded. This can result in disastrous consequences. These catastrophic events are brought about by human error when operators are placed in situations requiring cognitive resources beyond those currently available. Other system components are routinely monitored for their state of health. If deficiencies are found then corrective actions are taken or the mission is aborted. Similar monitoring and corrections are needed for the human component.

The goal of this report is to assemble pertinent information concerning the factors that produce suboptimal performance in human operators and the methods that can be used to detect the presence of these factors. Typically, these factors are considered in isolation. By bringing this information together in one report, decision makers and scientists will be able to consider the numerous factors that have deleterious effects on operator performance and can take measures to prevent catastrophic errors.

In this report, theoretical issues are presented as a framework for the discussions of the risk factors that reduce the functioning of human operators and the assessment methods for measuring these effects. The obstacles to implementation of operator functional state assessment in the "real world" are discussed. The demands of the work place are much more rigorous than those of the laboratory. For implementation in the operational environment, solutions for problems having a negative impact upon the operator and overall system performance must be demonstrated to be robust and repeatable. Without such qualities, operator functional state assessment will not be built into systems by managers and system designers nor will operators use it.

This report provides a comprehensive survey of the factors that negatively impact the operator's functional state to perform the job. These factors include environmental factors such as noise, acceleration and thermal stress. States within the individual operator can interfere with optimal performance and include illness, sleep loss and disruption of circadian rhythms. Task characteristics can also be problematic and include the cognitive and physical demands of the task.

Methods which can detect these effects are described. Identifying suboptimal operator states makes it possible to take corrective actions. The methods include physiological, performance, and subjective assessment. The rationale for each measure is presented as are the procedures required to make the measurements. The information provided by the measures is described as are the limitations and equipment required. Matrices are presented that can be used to determine which assessment method is appropriate for each of the risk factors that impair operator performance. Modelling and mathematical tools for data analysis are also presented.

L’évaluation de l’aptitude opérationnelle de l’opérateur humain

(RTO-TR-HFM-104)

Synthèse

L’opérateur humain est un élément décisif des systèmes complexes modernes. La complexité de ces systèmes, la rapidité du déroulement des opérations militaires modernes et la compression des effectifs, sont autant de facteurs qui contribuent aux sollicitations cognitives importantes subies par le personnel militaire. Malheureusement, ces systèmes ne tiennent pas compte de l’aptitude opérationnelle de l’opérateur humain. Le flux de l’information, le nombre de décisions et d’actions qui sont à prendre, peuvent dépasser la capacité cognitive de l’opérateur. Les conséquences peuvent en être désastreuses. Ces incidents catastrophiques sont le résultat des erreurs humaines qui se produisent lorsque des opérateurs sont mis dans des situations requérant des moyens cognitifs supérieurs à ceux dont ils disposent. L’intégrité des autres éléments constitutifs du système est contrôlée en permanence. En cas d’anomalie, soit des mesures correctives sont prises, soit la mission est abandonnée. Il serait souhaitable de doter l’élément humain d’une capacité similaire de contrôle et de remise en état.

Ce rapport a pour objectif de rassembler des informations pertinentes concernant les facteurs qui provoquent des performances sous-optimales chez l’opérateur humain et les méthodes permettant de détecter leur présence. Typiquement, ces facteurs sont considérés en situation isolée. L’incorporation de toutes ces informations dans un seul rapport permettra aux scientifiques et aux décideurs de considérer les différents facteurs ayant des effets nuisibles sur les performances des opérateurs et de prendre les mesures nécessaires afin d’éviter des erreurs catastrophiques.

Dans ce rapport, des questions de théorie sont présentées en tant que cadre pour la discussion des facteurs de risque qui nuisent au fonctionnement des opérateurs humains, ainsi que des méthodes d’évaluation permettant de les caractériser. Les obstacles à la mise en oeuvre des résultats de l’évaluation de l’aptitude opérationnelle dans « le monde réel » sont examinés. Les exigences du lieu de travail sont beaucoup plus rigoureuses que celles du laboratoire. Afin d’assurer leur mise en oeuvre dans un environnement opérationnel, les solutions de problèmes ayant un impact négatif sur les performances des opérateurs, ainsi que sur les performances globales des systèmes doivent être robustes et reproductibles. Sans cela, l’évaluation de l’aptitude opérationnelle de l’opérateur humain ne pourra pas être intégrée dans les systèmes par les responsables et concepteurs de systèmes et les opérateurs ne pourront pas l’exploiter.

Ce rapport présente un aperçu complet des facteurs ayant un impact négatif sur l’aptitude des opérateurs à exécuter leur travail. Ceux-ci comprennent des facteurs d’environnement tels que le bruit, les accélérations et les sollicitations thermiques. L’état physiologique d’un opérateur peut l’empêcher d’atteindre son niveau de performance optimal et peut inclure la maladie, le manque de sommeil et la perturbation des rythmes circadiens. Les caractéristiques des tâches imposées peuvent également être problématiques et inclure les exigences cognitives et physiques de la tâche.

Des méthodes permettant de détecter ces effets sont décrites. L’identification d’états physiologiques sous-optimaux permet de prendre des mesures correctives. Les méthodes comprennent l’évaluation physiologique, l’évaluation subjective et l’évaluation des performances. L’objectif de chaque mesure est présenté, comme les procédures nécessaires à sa réalisation. Les données résultants des mesures sont décrites, comme leurs limitations et le matériel nécessaire. Des matrices sont présentées, permettant de déterminer la méthode d’évaluation la plus appropriée pour chacun des facteurs de risque qui nuit aux performances des opérateurs. Des outils de modélisation et des outils mathématiques pour l’analyse des données sont également présentés.

Table of Contents

List of Figures

List of Tables

Human Factors and Medicine Panel

CHAIRMAN
Dr. Robert ANGUS
Director General
Defence R & D Canada – Suffield
PO Box 4000 – Station Main
Medicine Hat, Alberta T1A 8K6
CANADA

VICE-CHAIRMAN
Dr. Jean-Michel Clere
Chef du Département Sciences Médicales et
Facteurs Humains DGA/DSP/STTC/SH
8, Boulevard Victor
F-00303 Armées
FRANCE

Panel Executive

Col. Carel BANSE, MA
BP 25
92201 Neuilly-sur-Seine, FRANCE
Tel: +33 1 55 61 22 60/62 Fax: +33 1 55 61 22 98
Email: bansec@rta.nato.int or pelatd@rta.nato.int

Membership of Task Group "Operator Functional State Assessment" HFM-056/TG-008

Dr. Glenn WILSON, Chair
AFRL/HECP
2255 H Street
Wright-Patterson AFB, OH 45433-7022
UNITED STATES OF AMERICA
Email: glenn.wilson@wpafb.af.mil

CANADA
Mr. William FRASER
Aircrew Life Support Sector
Defence R & D Canada - Toronto
1133 Sheppard Ave West
P O Box 2000
Toronto, Ontario, M3M 3B9
Email: bill.fraser@drdc-rddc.gc.ca

FRANCE
Dr. Maurice BEAUMONT
Departement de Neurophysiologie
IMASSA
BP 73
E 91223 Brétigny Sur Orge CEDEX
Email: mbeaumont@imassa.fr

GERMANY
Dipl.-Ing. Morten GRANDT
FGAN e.V.
Neuenahrer Strasse 20
53343 Wachtberg-Werthhoven
Email: grandt@fgan.de

Dr. Alexander GUNDEL
DLR Institute of Aerospace Medicine
Linder Höhe
51147 Cologne
Email: alexander.gundel@dlr.de

LITHUANIA
Prof. Giedrius VARONECKAS
Institute of Psychophysiology & Rehabilitation
Vyduno Str. 4
Palanga 5720
Email: giedvar@ktl.mii.lt

THE NETHERLANDS
Mr. Hans VELTMAN
TNO Human Factors
Kampweg 5
P O Box 23
3769 ZG Soesterberg
Email: veltman@tm.tno.nl

SWEDEN
Dr. Erland SVENSSON
FOI
Department of Human Sciences
P O Box 1165
S-581 11 Linköping
Email: erland.svensson@foi.se

UKRAINE
Dr. Alexander. BUROV
All-Ukrainian Ergonomics Association
24/1 Nagorna Str., off.110b
04107 Kiev
Email: ab_is@ukrpost.net

UNITED KINGDOM
Prof. Bob HOCKEY
School of Psychology
University of Leeds
Email: bobh@psychology.leeds.ac.uk

Dr. Graham EDGAR
School of Health and Social Sciences
University of Gloucestershire
Francis Close Hall
Swindon Road
Gloucestershire GL50 4AZ
Email: gedgar@glos.ac.uk

Dr. Helen STONE
Formerly at:
BAE SYSTEMS Advanced Technology Centres-Sowerby
P.O. Box 5, FPC 267
FILTON, Bristol BS34 7QW

UNITED STATES
Dr. Thomas BALKIN
Department of Behavioral Biology
Walter Reed Army Institute of Research
503 Robert Grant Avenue
Silver Spring, Maryland 20910-7500
Email: thomas.balkin@na.amed.army.mil

Dr. Kirby GILLILAND - *Alternate*
Department of Psychology
University of Oklahoma
455 W. Lindsey Street, DHT705
Norman, OK 73019
Email: kirby@ou.edu

Dr. Robert E. SCHLEGEL, Secretary
School of Industrial Engineering
University of Oklahoma
202 W. Boyd, CEC 124
Norman, OK 73019
Email: schlegel@ou.edu

CDR Karl VAN ORDEN
Director - Science and Technology
Naval Health Research Center
P.O. Box 85122
San Diego, CA 92186-5122
Email: vanorden@nhrc.navy.mil

Non-Task Group Contributors

Mr. Gregory L. BELENKY
Department of Behavioral Biology
Walter Reed Army Institute of Research
503 Robert Grant Avenue
Silver Spring, Maryland 20910-7500
UNITED STATES OF AMERICA
Email: gregory.belenky@na.amed.army.mil

Mr. Robert CARTER
USARIEM
Kansas St., Bldg. 42
Natick, MA 01760-5007
UNITED STATES OF AMERICA
Email: robert.carter@na.amedd.army.mil

Mr. Samuel N. CHEUVRONT
USARIEM
Kansas St., Bldg. 42
Natick, MA 01760-5007
UNITED STATES OF AMERICA
Email: samuel.cheuvront@na.amedd.army.mil

Mr. Gary KAMIMORI
Department of Behavioral Biology
Walter Reed Army Institute of Research
503 Robert Grant Avenue
Silver Spring, Maryland 20910-7500
UNITED STATES OF AMERICA
Email: gary.kamimori@na.amed.army.mil

Ms. Mary KAUTZ
Department of Behavioral Biology
Walter Reed Army Institute of Research
503 Robert Grant Avenue
Silver Spring, Maryland 20910-7500
UNITED STATES OF AMERICA
Email: mary.kautz@na.amed.army.mil

Mr. Barry McGUINNESS
BAE SYSTEMS Advanced Technology Centres-Sowerby
P.O. Box 5, FPC 267
Filton, Bristol BS34 7QW
UNITED KINGDOM
Email: barry.mcguinness@baesystems.com

Mr. Daniel REDMOND
Department of Behavioral Biology
Walter Reed Army Institute of Research
503 Robert Grant Avenue
Silver Spring, Maryland 20910-7500
UNITED STATES OF AMERICA
Email: daniel.redmond@na.amedd.army.mil

Mr. Christopher RUSSELL
AFRL/HECP
2255 H Street
Wright-Patterson AFB, OH 45433-7022
UNITED STATES OF AMERICA
Email: christopher.russell@wpafb.af.mil

Dr. Michael N. SAWKA
Chief, Department of Thermal & Mountain
USARIEM, Kansas Street
Natick, MA 01760-5007
UNITED STATES OF AMERICA
Email: michael.sawka@na.amedd.army.mil

Ms. Nancy WESENSTEN
Department of Behavioral Biology
Walter Reed Army Institute of Research
503 Robert Grant Avenue
Silver Spring, Maryland 20910-7500
UNITED STATES OF AMERICA
Email: nancy.wesensten@na.amedd.army.mil

Chapter 1 – INTRODUCTION

The need for operator functional state (OFS) assessment is prompted by a growing worldwide concern with the consequences of performance breakdown by operators in safety-critical task environments. Military personnel more often must work with complex systems with increasing levels of automation. Despite (or because of) increasing automation, the human operator has an increasingly central role in the execution of tasks, in many cases made more difficult by increased mental workload (Wickens & Hollands, 1999). In addition, both work related and unrelated *risk factors* impose increased demands (such as G-acceleration, extreme temperatures, noise, sleep disturbances, stress, and time pressure) and present a cumulative challenge to stress adaptation mechanisms. Such issues have been generally appreciated for some time within the human factors community. Following the successful NATO ARW in Les Arcs, France (Hockey, Gaillard, & Coles, 1986), there have been a number of recent reviews of widely used methods for assessing workload, fatigue, and the impact of stress and task demands on performance and situation awareness (Backs & Boucsein, 2000; Hancock & Desmond, 2001). However, at the practitioner level, the analysis of performance breakdown has been hindered by the inadequacy of methods for taking account of the adaptive/compensatory behavior of human operators. It is now recognized that effective performance requires the operator to manage a trade-off between the benefits of maintaining primary task goals (requiring sustained effort) and the costs of depleting limited energetical resources – resulting in fatigue and reduced capacity for further task performance (Hockey, 1997). The need to preserve resources is essential if operators are to respond effectively to unexpected demands or emergency situations, such as an unanticipated navigational hazard or failure of a normally reliable automatic control system.

It is often not possible to tell whether an operator is capable of carrying out a task by simply examining overt performance because of the strategic reallocation of mental capacity. However, sophisticated analysis may reveal “latent decrements”, in the form of increased effort and strain, errors in (less critical) secondary tasks, or increased activation and disturbances in the physiological systems driving effort and task engagement. The same kind of adaptive mechanisms have been identified in the response to stress and difficult working conditions, such as cognitive load, noise, sleep deprivation and shift work. Skilled, highly-motivated operators in real-life, safety-critical tasks normally maintain overt performance very effectively, even under severe demand and stress. Where breakdown does occur, it is often characterized by a “graceful degradation” rather than catastrophic failure. For a period before manifest performance degradation can be observed, the operator is likely to be in a state of limited functional competence, being able only to manage predictable, routine task demands, or produce bursts of high-effort control. By monitoring the development of such states, serious consequences of performance breakdown may be prevented.

OFS assessment should enable the prediction of professional performance of a particular operator on a particular day. The assessment is likely to be part of a larger scale monitoring system, and to be based on a psychophysiological model. The goal of this report is to bring together in a single document a discussion of the stressors that affect operator performance and a listing of assessment methods that can be used to assess the operator’s functional state.

1.1 DEFINITIONS

1.1.1 Operator Functional State (OFS)

OFS is defined as the multidimensional pattern of human psychophysiological condition that mediates performance in relation to physiological and psychological costs. OFS results from the synthesis of operator characteristics, current operator condition, and the operator's interaction with operational requirements. Because of major inter-individual differences, as well as systematically different effects of task and environmental conditions, OFS measures can be understood only by reference to two other concepts – background state and baseline state.

1.1.1.1 Background State

The background state represents the averaged, unloaded (resting) state of the operator, shed of all responsibilities and goals. This can be taken as a kind of *state signature* and reflects a variety of psychological, physiological, and cognitive personality profiles. A specific vector of psychological and physiological values may indicate minimal loading for one individual and maximum stress for another individual. An individual's unstressed background state must be known in order to make meaningful statements about changes reflected in the individual's state under loading or stress conditions (based on the current vector of parameter values). Although it is expected that some aspects of the personality profile may exhibit small changes from day-to-day, in general the background state would be expected to be fairly stable.

1.1.1.2 Baseline State

An alternative to the background state is the operational baseline state. This is defined as the local, non-stressed state of the operator prior to being actively engaged in a task. The baseline state is a specific response to the prevailing conditions. While clearly related to the background state, baselines may be above or below background levels (background is, theoretically, the average of all baseline states) and are naturally influenced by prior work, temporary individual state factors, and ambient environments. The baseline state can be taken to reflect the most relevant operational baseline for interpreting the effects of further task and environmental stressors in the situation.

1.1.1.3 Operational State

The operational state represents the functional state of the operator while engaged in a particular task under specific operational conditions. OFS results from the interaction of the baseline state of the operator with task demands and environmental stressors. In applying the concept of OFS, the match between an individual's fundamental (baseline) state and the operational state is important. Some individuals may need to make drastic changes from their baseline state to perform a task successfully, whereas others may have the benefit of a more natural match between their baseline state and that required for performing a particular task. Also considered important is the individual's adaptability (the ease with which he or she is able to move between state levels without experiencing strain).

1.2 FRAMEWORK

OFS should be regarded as the result of many physiological and psychological processes that regulate brain and body in an attempt to maintain an individual in an optimal condition to meet the demands of the work environment (Gaillard & Kramer, 2000). Figure 1 provides a framework for operator state assessment in which important concepts related to OFS assessment are included. The most important reason to assess the operator state is to prevent a performance breakdown. However, there is no direct relation between state and performance. The model describes an operator dealing with different aspects of the environment. He or she has to *process relevant information* (tasks) in order achieve an adequate level of *performance*. An operator can only do so when his or her state fits the required state for that particular task, otherwise the level of performance and associated regulatory costs will not be optimal. The fit between the required and the actual state is a continuous, mostly unconscious process.

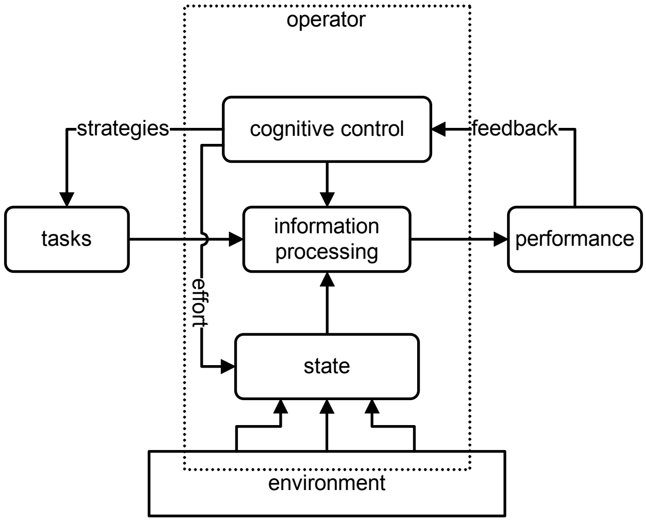

Figure 1: Conceptual Framework for Operator State Assessment.

An important mechanism for regulating large discrepancies is "mental effort." When operators are required to sustain performance on demanding tasks, and have a reduced baseline state because of a minor illness or sleep loss, or have to cope with external stressors (e.g., noise or high temperatures), they can only do so by increasing mental effort. This compensatory mechanism preserves performance levels but only at the expense of incurring additional costs. If this process of state regulation does not have the required result, or if there are too many costs involved in further effort investment, the operator can sometimes manage excessive task demand by changing the *strategy*. For example, he or she may decide to concentrate on the main tasks only and not to pay attention to less relevant tasks, or may reduce the reliance on immediate memory to control task input and make more use of external memory aids such as charts or tables. *Feedback* about performance is very important for this regulation process. The operator cannot adapt to the task requirements without adequate feedback.

Because of the "protective" (compensatory) effect of increased effort, it is clear that measuring performance is not sufficient to assess the state of the operator. The level of performance does not provide information about the costs involved in the adaptive response to stress. Particularly under conditions of performance protection (where there is no discernible breakdown under stress), physiological and subjective measures of OFS during task performance mainly reflect the amount of mental effort (strain) required to maintain task performance.

Continued investment of effort at a high level (strain) is uncomfortable, and the operator state is unstable. This means that performance is likely to break down if the state persists. A major challenge for the future is to be able to assess the state of the operator continuously and to be able to predict breakdowns in performance. One way of managing this is through adaptive automation (Scerbo, Freeman, & Mikulka, 2000). If the unstable state can be detected, demands on the operator may be reduced automatically by allocating more tasks to a computer or to other operators.

1.2.1 References

Backs, R.W., & Boucsein, W. (Eds.). (2000). *Engineering Psychophysiology: Issues and Applications*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Gaillard, A.W.K., & Kramer, A.F. (2000). Theoretical and Methodological Issues in Psychophysiological Research. In R.W. Backs & W. Boucsein (Eds.), *Engineering Psychophysiology: Issues and Applications* (pp. 31-58). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hancock, P.A., & Desmond, P.A. (2001). *Stress, Workload and Fatigue.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hockey, G.R.J. (1997). Compensatory Control in the Regulation of Human Performance under Stress and High Workload: A Cognitive-Energetical Framework. *Biological Psychology*, *45*, 73-93.

Hockey, G.R.J., Gaillard, A.W.K., & Coles, M.G.H. (1986). *Energetics and Human Information Processing*. Dordrecht, The Netherlands: Martinus Nijhoff.

Scerbo, M.W., Freeman, F.G., & Mikulka, P.J. (2000). A Biocybernetic System for Adaptive Automation. In R.W. Backs & W. Boucsein (Eds.), *Engineering Psychophysiology: Issues and Applications* (pp. 241-254). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Wickens, C.D., & Hollands, J.G. (1999). *Engineering Psychology and Human Performance* (3rd ed.). Prentice Hall.

1.3 LIMITATIONS IN THE APPLIED SETTING

There are numerous difficulties associated with implementing functional state assessment in the operational environment. The issues to overcome depend in part on the type of monitoring technologies that will be used as well as on the purpose of the monitoring. However, a number of problems are common to all implementations. Addressing these issues may require more effort than the development of the individual operator functional state technologies.

1.3.1 Operator Acceptance

Obtaining and retaining the acceptance of the individuals who will be monitored is both critical and difficult. The physical characteristics of the monitoring technologies will continue to be an issue for the wide acceptance of continuous functional state assessment. Sensors that require contact with the skin can in some instances be annoying or uncomfortable, possibly interfering in the performance of the tasks required of the operator. Equipment weight and volume, and additional cabling and connectors can also disrupt or annoy the operator, affecting the performance on the critical tasks. Non-contact optical and electromagnetic sensors that can monitor heart rate and brain activity, and miniaturization of electronics and computer technology can address these problems to some extent. Given the high workload of many operators, any additional training effort, or the imposition of additional tasks in preparing or maintaining the monitoring equipment will not be acceptable. However, if the monitoring and intervention can be demonstrated to clearly improve operator performance and thereby enhance overall system effectiveness, then operators and decision makers are likely to accept them. If the OFS monitoring significantly increases safety, then its use will possibly be mandated. The acceptance of anti-G suits by pilots in high performance aircraft is an example of added equipment that has been shown to have a definite utility in preventing G-LOC and saving lives. The suit and related equipment require that the pilots wear additional gear, and the aircraft must be equipped with sensors and other hardware. However, the added safety and mission enhancement promote its acceptance.

Once the physical problems are addressed, a major issue for wide acceptance of such monitoring is the real and/or apparent loss of privacy, (i.e., “big brother is watching”). Operator monitoring may also conflict with legal statutes, union agreements, or the historical culture of an organization. What is done with the data collected during monitoring and who has access may determine the degree of acceptance. Functional assessment of the operator to control life support systems or schedule rest periods may be acceptable, especially if the data are never stored, or are discarded after a limited time period. Analysis and reporting of data collected from populations may be acceptable; however, long-term tracking of individual performance, potentially resulting in disciplinary measures, will not be popular. In those cultures where performance monitoring is already acceptable (e.g., air traffic controllers), OFS monitoring may be welcomed. In other environments (e.g., aircraft piloting), monitoring may be regarded as obtrusive or threatening. Acceptance will vary across international boundaries as well. Clear statements regarding the privacy of the data, cooperation with unions, and implementation of security and encryption techniques in the transmission and storage of data will be critical.

1.3.2 Contextual Issues

The interpretation of physical and cognitive data in a real-world monitoring situation is more complex than in a laboratory or even a controlled field study. For this reason, data other than that obtained from the operator may be required, such as flight parameters, road conditions, weather conditions, and the operational situation. The context in which the physiological and cognitive data are collected is critical in interpreting and acting on the data. An example of this need for context would be when monitoring the heart rate of pilots. Increases in heart rate could be used to signal potential pilot mental overload. However, several normal flight conditions are associated with increased heart rate (i.e., take off and landing). If the context is not considered, then erroneous intervention would occur with potentially negative effects. Context consideration will help avoid this type of error. Even in this example, it would be necessary to establish activation limits for heart rate changes so that higher than expected heart rates during landing could be used to signal a difficult landing that may require some intervention.

1.3.3 Physical Considerations

In addition to the equipment volume and weight concerns of the operators, the physical impact of the monitoring hardware on vehicle function can be an issue in mobile systems such as aircraft and trucks. The provision of power, total power consumption, cooling requirements, and electrical interference (EMI) are problems that must be addressed. This will be less critical for ground-based stationary systems. Monitoring systems should be both low-cost and rugged. The hardware should be field upgradeable and require minimal maintenance. Because almost all monitoring technologies will be computer-based, a robust operating system and system architecture should be implemented that can be maintained and upgraded remotely via the communications system. Diagnostic software should be included to ensure data integrity and accuracy. Data storage, security (encryption), compression, and transmission must be considered critical, and the systems must be designed to have minimal impact on the communications bandwidth available for operational requirements. These are the same requirements that must be taken into account when any new equipment is added to a system. As previously discussed, if the end result clearly demonstrates an advantage, then the physical problems are likely to be addressed. Many of these considerations are routinely encountered and addressed when new equipment is added to existing systems (i.e., upgraded communication equipment replacing older, less capable equipment).

1.3.4 Data Mining

If one implements an operator assessment capability where the raw data are stored for off-line analysis as well as for research purposes, a wealth of new information will be obtained. This data mining capability will permit the validation of existing measures and the development of new measures. The volume of material will be many orders of magnitude greater than that acquired in a laboratory setting. This is primarily due to the longer duty hours and continuous nature of work for operators in their work environments. Advanced bioinformatic techniques have been developed to address the problems of data mining and analyzing the very large quantities of data generated in the human genomic and proteomic projects. Similar efforts have been undertaken for the financial sector. Standardized descriptors of datasets based on the Extensible Markup Language (XML) are widely used in a large variety of data-intensive research and commercial fields. The development of an XML variant specific to OFS assessment (i.e., OFSaXML) may be warranted. It will be necessary to develop the appropriate database and advanced processing software using distributed Web-based computer resources to automate the analytical techniques described in this report.

A major challenge is to continue the development and implementation of software and hardware systems that can provide the real-time data processing and automated interpretation required in operational settings. Real-time signal conditioning, waveform analysis, feature extraction, data reliability analysis, artifact detection and rejection, contextual analysis, data fusion, trend detection, and trend prediction may be some of the software components required.

1.3.5 Needed Validation Work

The validity of OFS measures must be demonstrated in the operational environment before they will be widely accepted and applied. While a number of investigations have produced data illustrating the relationships between various measures and OFS, many additional investigations are required. The operational world is highly complex and varied. Thus, the validity of OFS measures must be demonstrated in wide-ranging real-world situations. Furthermore, because of the extensive variations in the cognitive demands placed upon operators by the myriad of available jobs, the validity of the various measures in these situations must be demonstrated.

1.4 ADVANCED RESEARCH WORKSHOP

1.4.1 Purpose

An Advanced Research Workshop (ARW) was held in the Spring of 2002 (4-7 April) at Il Ciocco, near Lucca, Italy. The ARW, entitled *Operator Functional Status and Impaired Performance in Complex Work Environments*, was designed to broaden the framework of the RTO Task Group's terms of reference, and to maximize the expertise available for the compilation of this Report. ARW participants included approximately half of the Task Group membership and about thirty other contributors drawn from fourteen countries.

1.4.2 Process and Structure of the ARW

In order to maximize the interactive and discussion aspects of the meeting, there were only a small number of formal key papers, with other participants presenting brief position papers. The key papers were designed to draw together the leading data and theories in the field, and to act as a stimulus to the discussion. These papers addressed:

- Operator functional state and performance degradation – theoretical and methodological issues
- Evaluation of fatigue during and after work
- Operator functional state and pilot workload
- Sleep and work schedules as predictors of alertness and performance
- Heart rate variability in the evaluation of functional status during training
- Adaptive automation matched to human operator mental workload
- Functional status and regulatory processes in stress management
- Operator functional status and the prediction of fitness for duty (readiness to perform)
- Detecting low vigilance in operators through behavioral and physiological measures.

Towards the end of the ARW, one and a half days were devoted to panel discussions in three small groups. These groups focused on three aspects of OFS:

- Group A: Conceptual and theoretical foundations (Chair: Anna Leonova, Moscow, Russia)
- Group B: Methodological and assessment issues (Chair: Tony Gaillard, TNO Soesterberg, The Netherlands)
- Group C: Practical implications (Chair: Raja Parasuraman, Washington DC, USA).

There were a number of specific conclusions from these group discussions. Group A assumed the prevailing view of active state management as an option to protect performance standards under stress or extreme workload, with consequences (costs) for secondary performance, physiology, and subjective state. However, the group emphasized that OFS cannot be simply related to physiology, performance, or subjective measures, but must be linked explicitly to the work environment or task context. The pattern of the prevailing state interacts with the context to produce a performance decrement (or no decrement), either with some measured cost or no cost (if performance protection is not regarded as important by the individual).

Group B examined assessment methods within a highly interactive discussion, in which conflicts and diverse positions were evident. The group members agreed, however (as did Group A), that assessment must be considered within the work context. In particular, the time frame of measurement and prediction needs to be specified (next few minutes? next shift? sustained missions? even across the lifetime of the individual?) The group also agreed that the basis of any OFS analysis should be measurement at an individual level (i.e., referenced to a particular operator), thus highlighting the need for an individual database.

Group C primarily considered the application of OFS information to adaptive automation. A major application problem was how to trigger shifts in control between the operator and the system as a result of measured changes in state. A major issue in modeling adaptive automation was the hierarchy of intervening states between measured variables and triggered changes. Does the model need to recognize strain and fatigue, or simply act upon the non-specific state ‘alert’? How much does the model need to know about the dissociation between performance and strain/fatigue? The discussion also considered policies for using state information (e.g., whether triggered changes should be advisory or mandatory), and how these changes might be regulated and made acceptable.

1.4.3 Emergent Themes

A number of conclusions were common to all three groups, and other themes emerged in subsequent discussion within both the ARW and the Task Group. First, it was generally agreed that, in order to achieve progress, it is necessary to take an individual approach rather than rely on group norms. Second, more reliable assessment methods are needed, along with sufficient relevant data to make decisions at a practical level. A related point is that there is a major need to develop techniques for assessing patterns of responding rather than simply relying on individual measures. There is also a major need for studies to be conducted either in field situations or in high-fidelity simulations. Laboratory studies, while necessary to clarify cause-effect relationships, have difficulties associated with having participants engage in meaningless (or not very important) tasks. The research questions are often, as a result, examined in a naïve manner. For example, what is important is not workload *per se* but the internal effort that must be exerted when the task goals are taken seriously. The development of a “physiome” database (similar to a “genome”) is considered a major target for this work, allowing individual OFS management problems to be addressed effectively. There was also an expressed need to develop model scenarios that can be used across different stressors, thus allowing more effective comparison and generalization to other operations.

1.4.4 Output

The proceedings of the ARW will be published as an edited volume in the NATO Science Series by IOS Press, Amsterdam (Hockey, Gaillard, & Burov, 2003).

1.4.4.1 References

Hockey, G.R.J., Gaillard, A.W.K., & Burov, A.Yu. (in press). *Operator Functional State and Impaired Performance in Complex Work*. NATO ASI Series, Series A, Life Sciences, New York: Plenum.

Chapter 2 – TASK GROUP APPROACH

2.1 FUNCTIONAL STATE DIMENSIONS

2.1.1 Risk Factors

Three main risk factors are identified that affect the operator state and (as a consequence) performance: baseline individual state, task characteristics, and environmental factors. Individual state refers to the aspects within an individual (e.g., illness, circadian rhythm, and fatigue). Task characteristics refer to the physical and mental aspects of the task (e.g., load, action requirements, and monitoring demand). Environmental factors refer to ambient external conditions (noise, heat, and social factors).

2.1.2 Assessment Methods

The simplified model in Figure 1 shows that there is no direct relation between operator state and performance. Assessment of OFS requires measurement of performance, physiological state, and subjective reports – both state changes and task strategies. Furthermore, operator state is a multi-dimensional concept, and different aspects cannot always be measured directly. In some circumstances it is enough to have some global information about state; in other situations, more detailed information is necessary to predict effects of performance. It may also be important to track state changes over time in order to more effectively understand the nature of the adaptive processes involved. The consequence of this is that there is no general measure or even standard set of measures that can be used to assess operator state *per se*. The choice of measures should depend upon which aspects of state are most relevant to the current operational condition – the task, the person, the environment, and the circumstances in which the work is being done. The available measures that will be outlined in this document can be categorized into three groups: *physiological, performance,* and *subjective* measures. Measures from these categories provide different information about operator functional state. The sensitivity and applicability of many measures within these categories are described in the subsequent chapters of this document.

2.2 DOCUMENT STRUCTURE

2.2.1 Ranking Scheme

In order to facilitate the use of this document, matrices were developed to permit readers to quickly locate the measures that are suggested for assessing each risk factor. In each matrix the risk factors are listed as column headings and the assessment methods are listed as row headings. The risk factors are grouped as *Environmental Factors* (Table 1), *Task Characteristics* (Table 2), and *Individual State* (Table 3). The rows are grouped as *Physiological Measures* or *Subjective Measures*. Performance measures are not listed because their use typically depends upon the particular situation, and many of the performance tests could be used for most of the risk factors. Each cell is either blank or contains a number and a letter. Blank cells represent the situation where a particular measure was not considered appropriate for that risk factor. The number in a cell represents a scale from 1 to 3 indicating the measure's validity and utility for that risk factor. A *1* denotes a validated measure and one that is felt to be the "gold standard" for that risk factor. A *2* signifies that this is a useful measure, which is perhaps not validated or not considered the "gold standard" for that risk factor. A *3* represents an assessment measure that has been used in some applications and may also serve a niche function. The letter following the number indicates the estimated ease of use for each assessment method. An *A* designates an assessment method that is easy to use and is readily applied. A *B* designation indicates that the measure is deemed difficult to use but may have a promising future.

Table 1: A Matrix of the Environmental Factors and Assessment Methods presented in this Report. See the above text for an explanation of the nomenclature.

		ENVIRONMENTAL FACTORS					
		Accel -G	Drugs	Hyperbaric	Hypoxia	Noise & Vib	Thermal
PHYSIOLOGICAL MEASURES							
Actigraphy	Actigraphy						
Blood Flow		1B			2B		
BP	Blood Pressure	1B	2A			3A	2A
Core Temp	Core Temperature						1A
ECG (HR,HRV)	Electrocardiography	1A	1A	1A	2A?	2A	1A
EDA	Electrodermal Activity					2A	1B?
EEG	Electroencephalography	2A	1A	1A	1A		
EMG	Electromyography	2A	3A				
EOG/EyeMov	Electrooculography	3A	2A				
fMRI/imaging	Functional MRI		3B		3B		
Hormonal	Hormonal	3B	2A	1A	2A		
NIRS	Near-Infrared Spectro	2A			1A		
Oximetry	Oxygen Measurement	1A			1A		
Respiration	Respiration Parameters		2A	1B	1A		
SUBJECTIVE MEASURES							
NASA TLX	NASA Task Load Index						
Fatigue Scale	Brooks-Samn Perelli	1A	1A	1A	2A	1A	1A
POMS (Mood)	Profile of Mood States	2A	1A	2A	1A	1A	
SS	Sleepiness Scales						
Sleep Diaries	Sleep Diaries		1A				

Table 2: A Matrix of the Individual Risk Factors and Assessment Methods presented in this Report. See the above text for an explanation of the nomenclature.

		INDIVIDUAL RISK FACTORS				
		Circadian	Hydration	Illness	Mental Fatigue	Sleep Loss
PHYSIOLOGICAL MEASURES						
Actigraphy	Actigraphy	2A		3A	3A	1A
Blood Flow						
BP	Blood Pressure	2A				1A
Core Temp	Core Temperature	1A	1B	3A		
ECG (HR,HRV)	Electrocardiography	2A	2A	3A	3A	2A
EDA	Electrodermal Activity	?			2A?	1A
EEG	Electroencephalography					
EMG	Electromyography					
EOG/EyeMov	Electrooculography					
fMRI/imaging	Functional MRI	2A	2A	3A		
Hormonal	Hormonal	1A				
NIRS	Near-Infrared Spectro		2B			
Oximetry	Oxygen Measurement		3B			2A?
Respiration	Respiration Parameters				3B	3B
SUBJECTIVE MEASURES						
NASA TLX	NASA Task Load Index				2A	
Fatigue Scale	Brooks-Samn Perelli	2A		3A	1A	1A
POMS (Mood)	Profile of Mood States	2A		3A	1A	1A
SS	Sleepiness Scales	1A		3A	1A	1A
Sleep Diaries	Sleep Diaries	1A				1A

Table 3: A Matrix of the Task Characteristics Risk Factors and Assessment Methods presented in this Report. See the above text for an explanation of the nomenclature.

		TASK CHARACTERISTICS	
		Physical Load	Cognitive Load
PHYSIOLOGICAL MEASURES			
Actigraphy	Actigraphy		2A
Blood Flow		1A	2A
BP	Blood Pressure		
Core Temp	Core Temperature	3A	
ECG (HR,HRV)	Electrocardiography	1A	2A
EDA	Electrodermal Activity		2A
EEG	Electroencephalography	1A	
EMG	Electromyography	2B	
EOG/EyeMov	Electrooculography	2B	
fMRI/imaging	Functional MRI	1A	
Hormonal	Hormonal	2A	
NIRS	Near-Infrared Spectro	2A	
Oximetry	Oxygen Measurement		
Respiration	Respiration Parameters		2B
SUBJECTIVE MEASURES			
NASA TLX	NASA Task Load Index	2A	1A
Fatigue Scale	Brooks-Samn Perelli	1A	1A
POMS (Mood)	Profile of Mood States		
SS	Sleepiness Scales		
Sleep Diaries	Sleep Diaries		

2.2.2 Document Navigation

The risk factors and assessment methods are arranged alphabetically within each category in the body of this document. Use the above matrix to locate the risk factor of interest, then locate the recommended assessment method. If using the matrix on a computer, you should be able to click on the column heading of interest to automatically be taken to that section of the report. By clicking on the row heading you will automatically be taken to the section on that assessment method. In addition, you can locate the risk factors and assessment methods using the table of contents at the beginning of this report.

Authorship of the sections of this document are indicated in the table of contents. The author(s) of the sections are listed following the section name. Further information about each section can be obtained by using the contact information for the authors listed in the Membership of Task Group and Non-task Group Contributors.

Chapter 3 – RISK FACTORS

3.1 ENVIRONMENTAL

3.1.1 Hyperbaric Environments

3.1.1.1 Definition and Measurement

A hyperbaric environment refers to the exposure of divers to greater than normal atmospheric pressure. With the increased atmospheric pressure, the partial pressure of oxygen, nitrogen, and helium gas in the breathing systems is increased, resulting in a greater concentration of dissolved gases in a diver's tissues. Elevated oxygen partial pressure can result in the damage of lung and nervous system tissue, with subsequent seizures and death. The elevated concentration of nitrogen in brain cell membranes leads to nitrogen narcosis, resulting in mild to severe performance decrements. Helium gas, used as a replacement for nitrogen during very deep diving, can also have adverse effects leading to hyperactivity of the nervous system. Divers are typically exposed to the additional stresses of cold, anxiety, and intense physical workload. A detailed overview of performance issues in hyperbaric and diving environments is provided by Adolfson and Berghage (1974).

3.1.1.2 Background

Nitrogen narcosis is the most commonly observed reason for diver performance degradation, observable to some degree in almost all dives to greater than 30 m depth. The first reported cases of compressed air narcosis go back as far as 1835 (Bennett, 1982). Nitrogen narcosis causes feelings of euphoria and intoxication, but recovery is essentially instantaneous when divers ascend to shallower depths.

3.1.1.3 Effect on Performance

Numerous studies have quantified the impact of nitrogen narcosis on various aspects of performance, including arithmetic, reaction time, logical reasoning, and standing postural steadiness. Bennett (1982) provides a comprehensive review of nitrogen narcosis and an extensive list of references. A consensus conclusion of all these studies is that there is an increasing decrement in performance with increasing depth.

3.1.1.4 Assessment Methods

Cognitive performance tests, EEG, and evoked response techniques can be used to quantify the degree of performance impairment with nitrogen narcosis in compression chambers. There is a high correlation between the results from cognitive and electrophysiological assessment techniques (Bennett, 1982). However, the use of these electrophysiological techniques for routine research or operational monitoring during actual dives remains problematic. If voice communication is available, automated speech analysis software may provide a more practical method for monitoring the cognitive state of a diver.

3.1.1.5 References

Adolfson, J., & Berghage, T. (1974). *Perception and Performance Under Water*. New York: John Wiley & Sons.

Bennett, P. (1982). Inert Gas Narcosis. In P.B. Bennett & D.H. Elliott (Eds.). *The Physiology and Medicine of Diving* (3rd ed., pp. 239-261). London: Bailliere Tindal.

3.1.2 Hypoxia

3.1.2.1 Definition

Oxygen is one of the most important elements required for maintaining the normal functioning of living organisms. The absence of an adequate supply of oxygen to the living organism is termed *hypoxia*. Humans are extremely sensitive to and vulnerable to the effects of oxygen deprivation, and severe hypoxia results in the rapid deterioration of most bodily functions. Mental processes are sensitive to hypoxia, and even low levels of hypoxia result in the degradation of perceptual and cognitive functions.

Hypoxic hypoxia resulting from a reduction in the oxygen tension in inspired gas is a common form of oxygen shortage in general aviation and mountain climbing (*hypobaric hypoxia*). *Ischaemic hypoxia* is the consequence of a reduction in the blood flow through the tissues. Ischaemic hypoxia is caused by general circulatory failure as may occur after the drop in cardiac output and blood pressure associated with exposure to high sustained accelerations as well as rapid onset of G-load. *Hyperventilation* results in *hypocapnia*, a condition characterized by a reduction in the alveolar and arterial tensions of carbon dioxide. Hypocapnia is a normal concomitant of hypoxia, and both conditions produce almost identical symptoms (Ernsting, Sharp, & Harding, 1988).

3.1.2.2 Background

Numerous studies have investigated the effects of hypobaric hypoxia on mental performance. A reduction of approximately 25% in the partial pressure of oxygen in the atmosphere (associated with ascent to an altitude of about 2500 m) produces impairment in some aspects of mental performance. A sudden exposure to a rapid decompression, reducing the partial pressure of oxygen to about 10% of its sea level value, will cause unconsciousness within about 10 to 15 s. In the past, lack of oxygen in flight has killed many military aircrews, and many more crewmembers have experienced impaired performance due to hypoxia (Ernsting et al., 1988).

Modern military aircraft are highly maneuverable and capable of steep turns that produce rapid acceleration forces. These high G-forces cause ischaemic hypoxia, which produces effects such as tunnel vision and rapid Loss of Consciousness (LOC). The physiological effects of acceleration forces have been intensely studied for many years. The effects on a pilot's cognitive performance during high and sustained acceleration have, on the other hand, been studied to a lesser extent. An important area for future research is the combined effect of acceleration and mental load on pilot performance.

3.1.2.3 Effects on Performance

Psychomotor tasks such as simple reaction time are relatively unaffected up to altitudes of about 5000 m, although wide individual variability exists. However, more complex psychomotor tasks such as choice reaction time and pursuit or control tasks are more sensitive and are affected at lower altitudes. Psychomotor tasks are further compromised by the impairment of muscular coordination produced by moderate and severe hypoxia (Cheung & Hofer, 1999; Ernsting et al., 1988).

Cognitive tasks such as conceptual reasoning, short-term and long-term memory, paired word association, and mood become affected at an oxygen tension comparable to an altitude of about 4000 m. The severity of the decrement increases as a function of the difficulty and complexity of the task. Experience and training make the performance of pilots less vulnerable to hypoxia (Bartholomew et al., 1999; Du, Li, Zhuang, Wu, & Wang, 1999; Ernsting et al., 1988; Paul & Fraser, 1994; Shukitt-Hale, Banderet, & Lieberman, 1998). Acute exposures appear to have a larger negative impact on cognitive functioning than exposures over a longer period of time (Crowley et al., 1992). Nesthus, Rush, and Wreggit (1997) found that hypoxic pilots committed more procedural errors during the cruise, descent, and approach phases of flight from 3050 m and during descent and approach from 3813 m. Some studies present conflicting results. Gustafsson, Gennser, Oernhagen, and Derefeldt (1997) found no deterioration in cognitive performance as a function of different levels of normobaric hypoxia in closed spaces (submarines). Rather, performance improved with time as a result of learning, despite the reduction in oxygen level.

Some researchers have found positive effects on physical performance from *increased* oxygen levels (i.e., *hyperoxia*) above 21% (Petersen, Dreger, & Williams, 2000). Andersson, Berggren, Groenkvist, Magnusson, & Svensson (2002) found no effects on cognitive performance or mood after inhalation of 100% oxygen.

The mechanisms responsible for the effects of hypoxia on cognitive functioning are so far not well understood. However, it has recently been shown in studies of the location of impairment of brain function that acute hypoxia influences early and preprocessing stages of information processing (Beach & Fowler, 1998; Stivalet, Leifflen, Poquin, & Savourey, 2000; Qin, Ma, Ni, Fu, & Cheng, 2001). In several studies, brain function has been analyzed by means of different EEG-measures (e.g., Cheng, Ma, Ni, & Wang, 1999). New imaging techniques for analyses of brain activity, such as near infrared spectroscopy (NIRS) will be of value in studies of the location of functional impairment.

3.1.2.4 References

Andersson, J., Berggren, P., Groenkvist, M., Magnusson, S., & Svensson, E. (2002). Oxygen Saturation and Cognitive Performance. *Psychopharmacology*, *162*, 119-128.

Bartholomew, C.J., Jensen, W., Petros, T.V., Ferraro, F.R., Fire, K.M., Biberdorf, D., et al. (1999). The Effects of Moderate Levels of Simulated Altitude on Sustained Cognitive Performance. *Applied Human Science*, *18*, 175-179.

Beach, C., & Fowler, B. (1998). Evidence that the Slowing caused by Acute Hypoxia is Modality Dependent. *Aviation, Space, and Environmental Medicine*, *69*, 887-891.

Cheung, B., & Hofer, K. (1999). Degradation of Visual Pursuit during Sustained +3 Gz Acceleration. *Aviation, Space, and Environmental Medicine*, *70*, 451-458.

Cheng, H.W., Ma, R.S., Ni, H.Y., & Wang, X.B. (1999). Effects of Mild and Moderate Hypoxia on Event-Related Potentials and the Comparison of Different Stimulus Paradigms. *Space Medicine and Medical Engineering*, *12*, 23-27.

Crowley, J.S., Wesenten, N., Kamimori, G., Devine, J., Iwanyk, E., & Balkin, T. (1992). Effects of High Terrestrial Altitude and Supplemental Oxygen on Human Performance and Mood. *Aviation, Space, and Environmental Medicine*, *63*, 696-701.

Du, J.Y., Li, X.Y., Zhuang, Y., Wu, X.Y., & Wang, T. (1999). Effects of Acute Mild and Moderate Hypoxia on Human Short Memory. *Space Medicine and Medical Engineering*, *12*, 270-273.

Ernsting, J., Sharp, G.R., & Harding, R.M. (1988). Hypoxia and Hyperventilation. In J. Ernsting & P. King (Eds.), *Aviation Medicine* (pp. 545-590). London: Butterworths.

Gustafsson, C., Gennser, M., Oernhagen, H., & Derefeldt, G. (1997). Effects of Normobaric Hypoxic Confinement on Visual and Motor Performance. *Aviation, Space, and Environmental Medicine*, *68*, 985-992.

Nesthus, T., Rush, L., & Wreggit, S.S. (1997). *Effects of Mild Hypoxia on Pilot Performance at General Aviation Altitudes* (FAA-AM-97-09). Oklahoma City, Oklahoma: FAA, Civil Aeromedical Institute.

Paul, M.A., & Fraser, W.D. (1994). Performance during Mild Acute Hypoxia. *Aviation, Space, and Environmental Medicine*, *65*, 891-899.

Qin, Y., Ma, R., Ni, H., Fu, Z., & Cheng, H. (2001). Locating the Impairment of Human Cognitive Function during Hypoxia. *Space Medicine and Medical Engineering*, *14*, 218-220.

Shukitt-Hale, B., Banderet, L.E., & Lieberman, H.R. (1998). Elevation-Dependent Symptom, Mood, and Performance Changes produced by Exposure to Hypobaric Hypoxia. *International Journal of Aviation Psychology*, *8*, 319-334.

Stivalet, P., Leifflen, D., Poquin, D., & Savourey, G. (2000). Positive Expiratory Pressure as a Method for Preventing the Impairment of Attentional Processes by Hypoxia. *Ergonomics*, *43*, 474-485.

3.1.3 Noise and Vibration

3.1.3.1 Definitions and Measurement

Audible acoustic signals (sound waves) are transmitted by air with a speed of 330 m/s. If sounds are uncomfortable, undesired, and/or dangerous to the human, these sounds are defined to be noise. Vibrations are low-frequency mechanical waves that are transmitted by solids. Both sound waves and vibrations are typically measured by physical measurement devices and can be described by their amplitude, frequency, and sound pressure level.

3.1.3.2 Background

Legal regulation of limit values for noise differs between countries. Recommendations have been provided for different task demands as shown in Table 4 (e.g., Neumann & Timpe, 1976). It should be noted that these recommended values are based primarily on feelings of annoyance.

Table 4: Recommended Threshold Values for Auditory Noise with regard to Task Demands (Neumann & Timpe, 1976)

Type of Task	**Recommended equivalent continuous sound pressure level L_{eq} [dB(A)]**
Mental creative work	45
Mental schematic work	55
Supervisory tasks, operating machines - low demands - high demands	 65 55
Tasks for which speech understanding is essential	80

Vibration thresholds that differentiate between dangerous and non-dangerous vibrations have not been determined. However, there are threshold criteria curves that provide estimates of accelerations that can be endured (see Figure 2) while maintaining specified proficiency levels. Threshold values to meet comfort criteria can be obtained by dividing the given values by 3.15. Values to meet safety criteria can be obtained by multiplying the given values by 2.

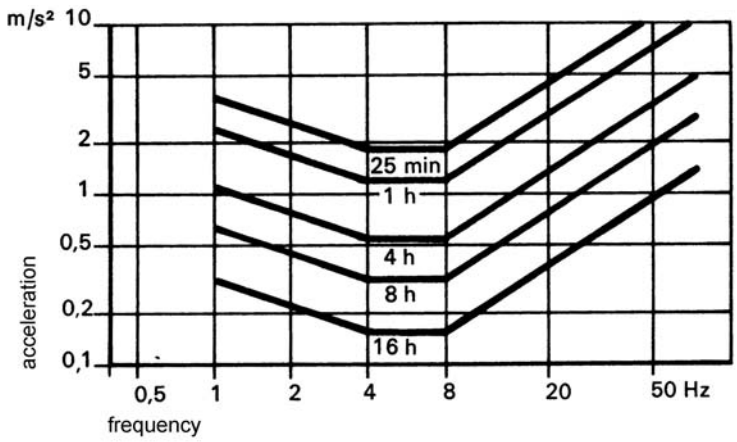

Figure 2: Threshold Values of ISO 2631 (Cited in Grandjean, 1991) for Combinations of Vibration Amplitude in the Vertical (-Z) Direction and Exposure Duration in order to Maintain the Ability to Perform Tasks (Fatigue-Decreased Proficiency Boundary).

3.1.3.3 Effects of Noise

In addition to the ill effects of noise on audition, it has been shown that noise may cause several extra-aural effects. To what extent these effects occur depends upon several properties of noise such as its intensity, its regularity (constant vs. intermittent), and its information content (understandable speech vs. mechanical sound). In the context of performance tasks, effects are also influenced by the nature of the working situation (e.g., whether a task is monotonous or stimulating, involves memory, attention or rapid decision making, or is perceived as high or low in priority; Jansen, 1970, 1989).

3.1.3.3.1 Disorders of Attention, Performance Degradation

Noise has minimal negative effects on the proficiency of physical work. With respect to mental work (i.e., work in which information processing plays an important role), it is obvious that noise disturbs verbal communication processes, especially the auditory perception of information.

Generally it is assumed that noise has negative effects on performance. However, it has been shown that under certain circumstances noise may have neutral or even positive effects on performance (e.g., if it superposes other disturbing noise with high information content).

More recently it has been shown (Hockey, 1997) that noise effects on performance can be compensated by an increase in effort under conditions of high motivation on the part of the operator. The cost of protecting performance may be manifested in increased psychophysiological activation, such as blood pressure and HRV suppression. Where individual motivation for a task is low, performance may not be protected in this way, and overt performance degradation is more likely to occur.

3.1.3.3.2 Disorders of Sleep

In this context, noise leads to reduced sleep duration, reduced deep sleep, longer periods of wake time, and longer time required to fall asleep.

3.1.3.3.3 Feelings of Annoyance

Whether noise leads to a feeling of being disturbed depends on the individual's mental attitude to the noise and the kind of noise. With respect to the first point, it is important to note that people who produce noise (e.g., by their machines or their work) are usually not disturbed by "their own" noise. In contrast, uninvolved persons are more disturbed by noise created by others. Other factors influencing the degree of annoyance were mentioned above and pertain to the noise characteristics. As with other psychological constructs such as intelligence or skill, annoyance must be understood as a complex structure of effects that may involve several psychological reactions (e.g., emotional workload, increase of activation, stress, or nervousness). Annoyance may result in degradation of concentration that is needed for cognitive tasks as well as for supervisory tasks.

3.1.3.4 Effects of Vibrations

Mechanical oscillations (i.e., vibrations) have significant ill effects on the human body whenever the frequencies of oscillation correspond to the resonance frequencies of body parts or organs (Table 5). Besides the physical effects, vibrations also affect psychophysiological and psychological reactions.

Table 5: Resonance (*Eigen*-) Frequencies of Human Body Parts with Respect to Different Postures and Oscillation Directions (Reference Coordinate System is X: Sagittal Axis, Y: Transverse Axis, Z: Vertical Axis of the Human Body standing in Upright Position; Dupuis & Hartung, 1989)

Posture	Body Part	Direction	Eigen-Frequency (Hz)
Lying	Feet	X	16 - 31
		Y	0.8 - 3
		Z	1 - 3
	Knee	X	4 - 8
	Ventral	X	4 - 8
		Y	0.8 - 4
		Z	1.5 - 6
	Chest	X	6 - 12
	Head	X	50 - 70
		Y	0.6 - 4
		Z	1 - 4
Standing	Knee	X	1 - 3
	Shoulder	X	1 - 2
	Head	X	1 - 2
	Whole Body	Z	4 - 7
Sitting	Trunk	Z	3 - 6
	Chest	Z	4 - 6
	Backbone	Z	3 - 5
	Shoulder	Z	2 - 6
	Stomach	Z	4 - 5 (7)
	Eye	Z	20 - 25

Vibrations are subjectively experienced as troublesome. The degree of discomfort primarily depends on the frequency, the acceleration, and the duration of exposure. The feeling of discomfort reflects the physiological effects and the resonance phenomena of various body parts.

With regard to psychophysiological reactions, it has been found that vibrations affect the cardiovascular and respiratory systems to only a minor degree. Much more important are the effects on the visual faculties. Vibrations with frequencies of about 4 Hz lead to a reduction in visual acuity. Degradations become extreme if frequencies are in the range of the resonance frequency of the bulbus (i.e., 20-25 Hz).

Due to the effects on the visual system, but also due to complications of motor information transfer, motor and sensorimotor task performance is degraded by vibrations.

3.1.3.5 Assessment Methods

3.1.3.5.1 Physiological Measures

Activation has an influence on the vegetative nervous system, which entails physiological reactions of interior organs. Various psychophysiological studies have shown that exposure to noise leads to an activation of the sympathetic system which may be reflected by an increase in blood pressure and heart rate, reduction in stroke volume, increase in pupil diameter, increase in metabolism, increase in muscle tension, dermal vasoconstriction, and reduction of digestive activity. Physiological measurement techniques that reflect the reactions mentioned above include electrocardiography, blood pressure, and electrodermal activity.

The factors that moderate the psychophysiological reactivity are the same as with psychological reactions (i.e., the type of noise, the individual sociological situation, and the attitude towards the noise and towards the task). If the noise contains a certain amount of information, it can be assumed that the same noise may lead to different intra-individual psychophysiological reactions and beyond these to significant inter-individual differences in psychophysiological reactivity. It has been shown that the physiological reactivity to noise is strongly related to the psychological stability of the individual. If the information content is low, physiological reactions in groups differing in physiological reactivity are essentially the same.

Noise-induced physiological effects are also affected by habituation. At low intensity, noise will lead to no physiological reaction. At higher intensity, an orientation reaction will occur at the beginning of the noise exposure. This reaction disappears if a person becomes accustomed to the noise. Above critical values of noise intensity and noise exposure time, no habituation occurs. In these cases, orientation reactions will change to defensive reactions.

Physiological measurement techniques that reflect reactions to the vibrations mentioned above during exposure and task performance are not known.

3.1.3.5.2 Subjective Measures

It is possible to explore psychological reactions to noise by the application of subjective techniques that reflect the psychological properties of affect and mood. An example of a technique in the English language is the Profile of Mood States (POMS) developed by McNair, Lorr, and Droppleman (1971). Similar techniques are available in other languages.

The exploration of psychological reactions to vibrations is also possible by applying subjective techniques that reflect the psychological properties of upset.

3.1.3.6 General Remarks

It has been noted above that there are significant differences in the psychological and psychophysiological noise reactivity both within and between subjects. Therefore, it seems to be questionable whether subjective assessment on the one hand or physiological measures on the other can be applied as meaningful methods for the assessment of mental degradation due to noise.

Since psychological and physiological techniques do not exclusively reflect reactions induced by noise or vibrations but also reactions due to task demands and other factors, the superior method is the direct (i.e., physical) measurement of noise and vibration in the workplace environment. However, with these techniques, aspects of comfort and annoyance cannot be evaluated.

3.1.3.7 References

Dupuis, H., & Hartung, E. (1989). Ganz-körper-schwingungen [Whole-Body-Vibrations]. In Bundesamt für Wehrtechnik und Beschaffung [German Federal Procurement Office] (Ed.), *Handbuch der ergonomie [Handbook of Ergonomics]*. Koblenz, Germany: BWB.

Grandjean, E. (1991). *Physiologische arbeitsgestaltung [Physiological Labor Organization]*. Thun, Switzerland: Ott Verlag.

Hockey, G.R.J. (1997). Compensatory Control in the Regulation of Human Performance under Stress and High Workload: A Cognitive-Energetical Framework. *Biological Psychology*, *45*, 73-93.

ISO 2631 (1974). *Guide for the Evaluation of Human Exposure to Whole-Body Vibration*. Geneva: International Standards Organization.

Jansen, G. (1970). Lärm im arbeitsraum [Noise in the Workplace]. In A. Mayer & B. Herwig (Eds.), *Betriebspsychologie [Industrial Psychology]*. Göttingen, Germany: Hogrefe.

Jansen, G. (1989). Physiologische und psychologische wirkungen von dauerlärm [Physiological and Psychological Effects of Continuous Noise]. In Bundesamt für Wehrtechnik und Beschaffung [German Federal Procurement Office] (Ed.), *Handbuch der ergonomie [Handbook of Ergonomics]*. Koblenz, Germany: BWB.

McNair, D.M., Lorr, M., & Droppleman, L.F. (1971). *Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.

Neumann, J., & Timpe, K.-P. (1976). *Psychologische arbeitsgestaltung [Psychological Labor Organization]*. Berlin: VEB Deutscher Verlag der Wissenschaften [German Publishing House of Sciences].

3.1.4 Pharmacological Mediators (Drugs and Medicines)

3.1.4.1 Background

In the modern operational environment, readiness to perform is largely determined by the operator's ability to perform mental work (e.g., the capacity to plan ahead, recognize and capitalize upon emergent opportunities, and effectively communicate and coordinate with other operators).

Among the many operationally relevant factors that can impact performance (e.g., anxiety, workload, time on task, distracting stimuli, and environmental temperature extremes) is the drug/medication status of the operator. In addition to substances of abuse (e.g., alcohol), there is the potential impact of drugs taken incidentally to treat both short-term and long-term health problems (e.g., antihistamines, anticonvulsants, and antihypertensives) as well as those substances that may be administered to protect against perceived theater-specific threats (e.g., vaccines against endemic diseases, antibiotics to combat biological warfare agents, and pyridostigmine to protect against nerve agents). A third category includes pharmacological agents that are administered for the express purpose of enhancing operational performance.

A comprehensive review of the performance effects of all drugs/medications that might be encountered in the operational environment would be a considerable undertaking – well beyond the scope of the present section. Instead, the focus here will be on the current status of pharmacological agents to enhance and sustain performance in the operational environment – specifically, stimulants to sustain performance when sleep is not possible and sleep inducers to enhance the recuperative values of sleep when its quality is compromised due to environmental or circadian factors (e.g., hypnotic drugs can be useful for inducing sleep during daytime rest periods and thus help maintain vigilance and performance overnight).

These pharmacological agents are of particular relevance because a primary, underlying physiological factor that delimits the brain's capacity to perform mental work is its sleep debt status. Although actual mental work output can be affected by a multitude of operationally relevant factors such as anxiety, workload, time on task, distracting stimuli, and environmental temperature extremes, the brain's range of effectiveness is ultimately a function of its physiological status. Of the various physiological insults likely to be encountered in the operational environment, sleep loss is the most common. This is because modern operations are increasingly continuous, 24-hour-per-day endeavors in which the opportunities for adequate sleep are seriously compromised. This is especially true of military operations, which are generally of two types:

- Continuous operations (CONOPS, commonly experienced by infantry), which take place over weeks or months, during which opportunities for obtaining adequate sleep are few (resulting in chronic sleep restriction), and
- Sustained operations (SUSOPS, commonly experienced by aircrew on long-range bombing missions), which are characterized by extended periods (i.e., 24+ hours) of total sleep deprivation.

Often, military operations are a mixture of both types: continuous operations punctuated by sustained operations.

Accordingly, there are two basic strategies available for pharmacologically enhancing alertness, and thus cognitive performance, in the operational environment:

1. Direct enhancement of alertness with stimulants when operational exigencies preclude recovery sleep (i.e., during sustained operations), and
2. Optimization of sleep with sleep inducers when the opportunity to obtain at least some sleep is available, but the ability to sleep is diminished by circadian and/or environmental factors (i.e., during continuous operations).

No country follows specific (and written) rules for the pharmacological management of the sleep/wakefulness cycle but, perhaps based on ethical considerations related to the perceived relative dangers associated with the administration of these drugs, some countries (such as France and the UK) seem to prefer the use of hypnotics (i.e., sleep inducers to manage daytime rest periods) rather than psychostimulants, which would be administered only when other options are not available or realistic. Both pharmacological strategies have been employed during military operations with apparent success – although the absence of proper scientific controls during actual operations has typically precluded scientific assessment of effectiveness.

3.1.4.2 Stimulants

There are many stimulant agents potentially available for use in the operational environment, including caffeine, *d*-amphetamine, modafinil, methylphenidate, pemoline, and nicotine, to name but a few. Of these, three are currently of particular interest:

1. Caffeine – because of its wide availability and status as a non-controlled substance (caffeine is already commonly used, albeit informally, to maintain alertness and performance in a wide range of operational environments).

2. Modafinil – because of claims that this substance may actually reduce the need for sleep).
3. *d*-amphetamine – the "gold standard" in terms of stimulant effectiveness, currently prescribed for a small number of operators (e.g., U.S. Air Force pilots, used only under strictly proscribed circumstances).

Although the performance-enhancing effects of *d*-amphetamine in the operational environment have been well established (Caldwell, Smythe, Leduc, & Caldwell, 2000), it is unlikely that its use will ever be widespread because of its high abuse potential. Therefore, the emphasis of this subsection is on the potential usefulness of caffeine and modafinil.

3.1.4.3 Caffeine

Caffeine is one of the most widely used drugs in the world (Dews, 1984a). It has been shown to have low toxicity and it produces no serious adverse physiological effects (Dews, 1984b). Caffeine is often used to counteract the performance and alertness deficits resulting from irregular work/rest schedules (Akerstedt & Ficca, 1997). Numerous studies have demonstrated that caffeine improves alertness and performance across night-time hours and during sleep deprivation (Bonnet & Arand, 1994; Smith, 1995). In addition, caffeine (32-600 mg) significantly decreases reaction times in auditory and visual choice reaction time tasks in non-sleep-deprived individuals (Babkoff, Mikulincer, Caspy, Carasso, & Sing, 1989; Babkoff, Sing, Thorne, Genser, & Hegge, 1989; Lorist & Snel, 1997; Smith, 1995), thus suggesting some non-specific performance enhancing properties.

In the military operational environment, fatigue and/or the loss of sleep are common and can lead to mission-threatening degradation of both physical and mental performance. Therefore, a safe, reliable, rapid, pharmacological means to reverse the performance and alertness degradation associated with fatigue/sleep loss is needed. Whereas other stimulants, like *d*-amphetamine, are also effective for counteracting sleep-loss-induced decrements in alertness and performance (Newcombe, Renton, Rautaharju, Spencer, & Montague, 1988), they are also more often associated with negative side effects or are generally recognized to have relatively greater abuse potential (Lagarde et al., 2000). By comparison, caffeine is a relatively safe, uncontrolled substance that is well tolerated, with few side effects (Dews, 1984a). It is also very effective. Virtually all previous studies have shown that caffeine reverses sleep-loss-induced performance, alertness, and mood deficits, even following prolonged (48-61 hours) wakefulness (Penetar et al., 1993). In a recently published monograph titled *Caffeine for the Sustainment of Mental Performance*, the Institute of Medicine (National Academies) concluded that caffeine is both safe and efficacious, and guidelines are suggested for its use to maintain alertness and performance in the operational environment (Institute of Medicine, 2001).

3.1.4.3.1 Mechanism of Action, Pharmacokinetics, and Side Effects

Caffeine is a potent central adenosine receptor antagonist that is commonly used as a stimulant to alleviate the effects of sleep deprivation. Pharmacokinetic parameters of caffeine solution are reported in Table 6.

Table 6: Pharmacokinetic Profile of Caffeine Solution (5 mg/kg; 350 mg) following Oral Administration in Normal Healthy Adults (Bonati et al., 1982)

C_{max} (µg/ml)	8.30 ± 0.10
T_{max} (h)	0.78 ± 0.10
$AUC_{0\text{-}24\,h}$ (µg/ml/h)	86.06 ± 11.5
$t_{1/2Ke}$ (h)	6.30

Caffeine is almost completely (99%) metabolized in the liver and thus classified as a low-clearance, flow-independent drug (Dews, 1984a). This means that its rate of deactivation is unaffected by delivery to the liver and can only be modified by a change in hepatic enzyme activity (Dews, 1984a; Ritschel, 1986; Welling, 1986). The pharmacokinetics of caffeine have been well documented both at rest (Bonati et al., 1982; Dews, 1984a) and under a variety of adverse conditions including high altitude (Kamimori, Eddington, et al., 1995) and sleep deprivation (Kamimori, Lugo, et al., 1995). Caffeine has been shown to exhibit dose-dependent pharmacokinetics. This means that its metabolism can be significantly slowed (saturated) when high doses (>500 mg) are administered (Kamimori, Eddington, et al., 1995; Kaplan et al., 1997).

In moderate doses (<300 mg), caffeine is well tolerated, producing few significant side effects (Robertson & Curatolo, 1984; Serafin, 1996). Even relatively high doses of caffeine (600 mg) are well tolerated by sleep-deprived individuals, with effects comparable to those reported in studies of non-sleep-deprived subjects using lower doses (Penetar et al., 1993). In addition, sleep debt status does not interact with caffeine to affect self-reports of side effects such as heart pounding, headache, sweating, and upset stomach. At doses of 300-350 mg, caffeine does not affect cardiac rhythm and rate, and does not cause clinically significant ventricular or supraventricular dysrhythmia (Newcombe et al., 1988).

3.1.4.3.2 Formulations

The route of administration can profoundly influence a drug's effects. Caffeine is most commonly ingested orally (i.e., in a beverage or capsule), so absorption occurs primarily in the stomach. Data from a recently completed study show that the absorption rate is significantly faster when caffeine is administered in chewing gum (which apparently facilitates its absorption through the oral mucosa) than when it is administered in a capsule (Kamimori et al., 2002). Performance data from this study have demonstrated that the onset of action (i.e., latency from caffeine administration to the appearance of measurable benefits to performance) is shorter following administration of caffeinated chewing gum than following administration of a capsule formulation (Kamimori et al., 2002). Data from a recently completed follow-on study have also demonstrated that vigilance can be effectively sustained across a single night of sleep loss with three 200 mg doses of caffeine (gum formulation) administered at 2-hour intervals (0300, 0500, 0700 hr). Recent evidence also suggests that alertness and performance can be sustained for extended periods by caffeine administered in a slow-release formulation. Such a formulation, at a single 300 mg dose, has been shown to maintain performance and vigilance during 13 h after ingestion without major side effects, due to its long effective half-life (Lagarde et al., 2000). These beneficial effects have been confirmed during a 36-h sleep deprivation with a single daily dose of 600 mg (Patat et al., 2000) and for a longer (64-h) continuous wakefulness period with 300 mg/dose given twice daily (Beaumont et al., 2001). Another interesting property of slow-release caffeine (300-mg) is that it can significantly shorten the recovery period following an eastbound flight across 7 time zones (Piérard et al., 2001). Thus, caffeine formulations can be tailored to optimize alertness and performance on the basis of anticipated operational needs.

3.1.4.3.3 Caffeine Status

In summary, prior studies have indicated that caffeine is well tolerated and has significant positive effects on alertness and cognitive performance over a wide range of doses, in both sleep-deprived and non-sleep-deprived individuals, both during the day and at night. It is safe, with few and mild side effects (except in individuals who may be especially sensitive to caffeine). Current efforts to develop a variety of caffeine formulations reflect a general appreciation of its safety and efficacy, and its potential usefulness in a wide variety of operational scenarios. It is anticipated that the usefulness of caffeine in the operational environment will be optimized when its effects have been quantified and modeled – promoting incorporation of caffeine dosing and timing recommendations into comprehensive systems for management of sleep/alertness in the field.

3.1.4.4 Modafinil

Modafinil (2-[(diphenyl-methyl)-sulfinyl]acetamide) is a novel synthetic stimulant currently available under the trade name Modiodal® in Europe and Provigil® in Great Britain and the United States. It is indicated for treating the excessive daytime sleepiness associated with the sleep disorder narcolepsy.

It has been claimed that modafinil enhances both subjective and objective alertness, improves performance, has low abuse potential, and produces none of the side effects associated with other stimulants such as amphetamine (Lyons & French, 1991). It has also been asserted that although it promotes alertness, it does not interfere with restorative sleep if the opportunity to sleep arises (Batéjat & Lagarde, 1999). Furthermore, it has been suggested that modafinil not only allows sleep recovery to occur without disturbance, but actually reduces the *need* for that recovery sleep (Buguet, Montmayeur, Pigeau, & Naitoh, 1995) – a claim that implies that modafinil does not simply *postpone* recovery sleep, but actually *replaces* recovery sleep (Pigeau, 2001). If these claims are substantiated, then modafinil clearly has potential applications in the operational environment.

3.1.4.4.1 Effects on Alertness and Performance

Few studies have been conducted to determine the effects of modafinil on cognitive performance and alertness during sleep deprivation in normal (i.e., non-narcoleptic) humans. Some scrutiny of the evidence to date is briefly described below.

In one of the first studies with normal humans, Pigeau et al. (1995) evaluated the effects of modafinil 300 mg, amphetamine 20 mg, or placebo administered three times across 64 hours of total sleep deprivation in 39 healthy male (n=38) and female (n=1) Canadian reservists. Drug (or placebo) was administered at 17.5, 47.5, and 57.5 hours of sleep deprivation (the latter to evaluate drug effects on recovery sleep, reviewed below). Compared to placebo, modafinil 300 mg improved performance as measured by correct responses per minute for four-choice serial reaction time, logical reasoning, and digit-span tasks. Modafinil significantly improved performance for nine hours after administration at 17.5 hrs of sleep deprivation, and for six hours after administration at 47.5 hrs of sleep deprivation. Similar results were found for a group administered amphetamine 20 mg, except that amphetamine improved performance for eight hours after the administration at 47.5 hrs. Thus, these results suggest approximately comparable efficacy between modafinil and *d*-amphetamine at the tested doses.

In a more recent study, Brun et al. (1998) evaluated the effects of modafinil on cognitive performance as well as core body temperature, plasma melatonin, cortisol, and growth hormone rhythms across 36 hours of wakefulness in eight healthy male subjects. Modafinil 300 mg or placebo was administered at 2200 Day 1 and 0800 Day 2 corresponding to 15 and 24 hours of continuous wakefulness. Performance on reaction time (RT – key press to a number on a computer screen) and "grammatical reasoning" (GR – comparison of a sequence of symbols to a reference statement) was evaluated every three hours. No main effect of sleep deprivation period on RT performance was reported. Although a significant drug condition by sleep deprivation period interaction was reported, not enough details were provided to determine whether this effect was due to modafinil versus placebo, or at which post-drug time points modafinil differed from placebo. However, visual inspection of mean response times for the GR test suggested that modafinil improved response time and suppressed the early morning drop in performance seen during sleep deprivation.

In a comprehensive study of seven healthy male Air Force participants, Lagarde, Batéjat, Van Beers, Sarafian, & Pradella (1995) evaluated the effects of modafinil 200 mg vs. placebo administered three times daily (1400, 2200, and 0600) on performance and alertness across 60 hours of total sleep deprivation. Following normal nocturnal sleep, subjects were awakened at 0700 and the first drug administration occurred at 2200 on Day 1. Testing continued until Day 3, with the final drug administration occurring at 1400 on Day 3. Performance measures included reaction time (RT), mathematical processing (MP), memory search (MS), spatial processing (SP), unstable tracking (UT), grammatical reasoning (GR), and a concurrent tracking/memory search (TMS) task. The entire task battery required approximately 40 min to complete. Tasks were evaluated for response time and percent errors. Lagarde et al. (1995) reported that, compared to placebo, modafinil improved *average* performance across the 60 hours of sleep deprivation. However, specific comparisons between modafinil and placebo at each test session for each task and dependent measure revealed relatively few statistically significant differences using a repeated-measures analysis of variance. Most statistically significant comparisons occurred after the second night of sleep deprivation, and these differences were primarily for a dependent variable called "deviation index" calculated for the "Unstable Tracking" and "Tracking/Memory Search" tasks. Few comparisons for response time on the Reaction Time task were significant, which was unexpected since response time on RT tasks has been shown to be very sensitive to sleep deprivation and to stimulant drugs (Penetar et al., 1994). The failure to find more robust differences between modafinil and placebo may have been due to the small sample size in this study (N=7). However, Lagarde et al. (1995) did not report means for test sessions in which differences between modafinil and placebo did not achieve statistical significance; thus, the extent to which modafinil and placebo conditions differed (albeit non-significantly) during these sessions is not known.

Lagarde et al. (1995) also studied the effects of modafinil 200 mg administered three times daily (1400, 2200, and 0600 hours) on objectively measured sleep latency (a measure of sleepiness) across 60 hours of sleep deprivation. Sleep latency tests were administered each day at 0300, 0900, 1400, 1700, and 2200 hours. The test times at 0300, 0900, and 1400 hours corresponded to 5, 3, and 8 hours since the last drug administration. (Note: although drug was administered at 1400, this administration would not contribute to drug-induced effects at 1400 since appreciable amounts of the drug would not yet have been absorbed). The test at 1700 corresponded to 3 hours post-drug and the test at 2200 hours corresponded to 8 hours post-drug. Modafinil significantly increased sleep latency (i.e., reduced sleepiness) compared to placebo at 0300, 1700, and 2200 hours on Day 2 (corresponding to 20, 34, and 39 hours of sleep deprivation), and at 0300, 0900, 1400, and 1700 hours on Day 3 (corresponding to 44, 50, 55, and 58 hours of sleep deprivation). However, the modafinil-mediated improvements in alertness were modest (sleep latencies of 3-4 min with modafinil vs. 1 min for the placebo group) although statistically significant. [In clinical settings (e.g., when screening for a sleep disorder), a sleep latency below 5 minutes indicates pathological sleepiness. Thus, although modafinil increased sleep latencies on Day 3 relative to placebo, modafinil did not increase sleep latency above pathological values.]

Caldwell, Caldwell, Smythe, & Hall (2000) tested the efficacy of modafinil for sustaining helicopter pilot performance (measured in a simulator), EEG activity, and mood (POMS) in a double-blind crossover design comparing three 200 mg doses of modafinil to placebo over 40 hours of continuous wakefulness. Although they found modafinil to be effective relative to placebo for maintaining pilot performance on 4 of 6 flight maneuvers, they also reported an increased incidence of vertigo, nausea, and dizziness – a side effect that could (if substantiated though replication) preclude the use of modafinil as a fatigue countermeasure in aircrew.

Most recently, Wesensten, Balkin, & Belenky (1999) conducted a study to compare the efficacy of three dose levels of modafinil (100, 200, 400 mg), a high dose of caffeine (600 mg), and placebo for restoring performance and alertness following 41.5 hours of continuous wakefulness. They found that psychomotor vigilance test (PVT) performance and alertness (as measured with the Maintenance of Wakefulness Test; MWT) were significantly improved by modafinil 200 and 400 mg relative to placebo, and effects were comparable to those obtained with caffeine 600 mg. Although a trend toward better performance at higher modafinil doses suggested a dose-dependent effect, differences between modafinil doses were not statistically significant. Performance enhancing effects were especially salient during the circadian nadir (0600 through 1000 hours). Few instances of adverse subjective side effects (nausea, heart pounding) were reported. Therefore, in this study it was concluded that modafinil was effective (but not significantly more effective than high-dose caffeine) for restoring performance and alertness during moderate sleep deprivation.

In addition, there is some preliminary evidence of an enhancing effect of modafinil on working memory in mice, but further studies are needed to confirm an effect in humans (Béracochéa et al., 2001).

It should be noted that although modafinil has been found to enhance both objective and subjective alertness in all published studies to date, it has also been reported to distort subjective self-estimates of performance capacity (Baranski & Pigeau, 1997), producing an "overconfidence" not evident in subjects who were administered *d*-amphetamine or placebo.

One possible explanation for this overconfidence effect is that sleep-deprived subjects have recourse only to temporal memory for making their confidence assessment, a temporal memory that is impaired due to prefrontal cortex fatigue. Since, at this point, it does not appear that modafinil is particularly effective for alleviation of prefrontal cortex fatigue, tasks involving the prefrontal cortex (e.g., tasks involving speech, divergent thinking, and temporal memory) might not be expected to show much beneficial effect of modafinil. In contrast, performance on these tasks is sustained by *d*-amphetamine, which acts more generally as a CNS stimulant (Pigeau, 2001).

It should also be noted that the overconfidence effect was observed with a single 300 mg dosage of modafinil. Using the same protocol for gathering confidence assessments, Baranski and colleagues did not observe a similar effect with 100 mg modafinil administered thrice daily (Pigeau, 2001). Therefore, perhaps the overconfidence effect does not appear until high doses of modafinil (>300 mg) are used. Nevertheless, such an effect could be undesirable in some operational contexts, and these findings suggest that further studies of the effects of modafinil on this and other aspects of judgment, more broadly defined, might be advisable.

3.1.4.4.2 Effects on Recovery Sleep

The claim that modafinil actually reduces the need for recovery sleep in humans following sleep deprivation is based primarily on a study by Buguet et al. (1995) in which 37 subjects were administered either modafinil, *d*-amphetamine, or placebo at three time points during 64 hours of continuous wakefulness, with the last administration occurring shortly before the first of two nights of recovery sleep. Based on the findings that (a) *d*-amphetamine produced strong, negative effects on several aspects of recovery sleep; (b) the placebo group showed the expected slow-wave sleep and REM sleep "rebound" effects; and (c) the sleep architecture of the modafinil group more closely resembled that of the placebo group than that of the *d*-amphetamine group, it was concluded that modafinil does not impair the ability to obtain recovery sleep. Based on the findings that the modafinil group spent less time in bed and showed a reduced total sleep time (TST) compared to the other groups, it was also concluded that modafinil actually reduced the need for recovery sleep relative to placebo and *d*-amphetamine. Although this is a tantalizing hypothesis, further studies will be required to determine whether the relatively reduced TST during recovery truly represents a decreased pressure to sleep, or merely reflects a residual alerting effect of modafinil.

3.1.4.4.3 Mechanism of Action, Pharmacokinetics, and Side Effects

Modafinil is currently thought to promote alertness primarily through inhibition of the dopamine reuptake transporter (Wisor et al., 2001). The human pharmacokinetic properties of modafinil following a single oral administration are summarized below in Table 7. Following oral administration, bioavailability of modafinil is nearly 100%. Moachon, Kanmacher, Clenet, & Matinier (1996) reported that peak plasma concentrations of modafinil are achieved 2-4 hours following a 200 mg oral dose (t_{max}). Modafinil's kinetics are linear as demonstrated in studies using dose ranges of 50-400 mg and 200-600 mg. It is extensively metabolized with less than 10% of the administered dose being excreted unchanged.

Modafinil is metabolized mainly via the CY-P450 enzyme CYP3A4. Its major metabolites are modafinil acid (CRL 40467) and modafinil sulfone (CRL 41056), and the main route of elimination is through urine. Modafinil acid (the main metabolite) is pharmacologically inactive. However, modafinil sulfone is pharmacologically active, with a half-life of approximately 12 hours.

Table 7: Pharmacokinetic Profile of Modafinil following Oral Administration in Normal Healthy Adults

Elimination half-life (h)	13.60 ± 0.81
C_{max} (mg/L)	3.73 ± 0.25
C_{min} (mg/L)	1.49 ± 0.17
t_{max} (h)	2.92 ± 0.30
AUC $_{0\text{-}12h}$ (mg/L/h)	33.85 ± 2.88

The modafinil product monograph (Cephalon UK, January, 1998) contains an extensive review of the safety information for modafinil based on over 2000 subjects. Highlights of the monograph are summarized here.

The main subjective side effect reported with the use of modafinil is headache. In a large multicenter study, a number of subjects reported that headache increased in a non-dose-dependent fashion (0 mg = 36%; 200 mg = 52%, 400 mg = 51%). Severity of reported headaches was mostly mild to moderate. In another multicenter study, the percentages of subjects reporting headaches were 44% for placebo, 42% for 200 mg, and 54% for 400 mg. The most common adverse events reported by subjects taking the highest dose of modafinil (400 mg) in these two multicenter studies are listed in Table 8.

Table 8: Adverse Events Reported by Sleep Disorder Patients taking Oral Modafinil 400 mg

Body System	Adverse Events (listed in order of frequency)
Body as a whole	headache, hypothermia, infection, back pain, pain, abdominal pain, fever
Digestive	nausea, diarrhea, dry mouth, anorexia, dyspepsia
Respiratory	rhinitis, pharyngitis, lung disorder, increased cough
Nervous	nervousness, dizziness, anxiety, depression, cataplexy, insomnia
Musculo-skeletal	myalgia
Urogenital	dysmenorrhea
Skin and appendages	rash
Haemic and lymphatic	eosinophilia

In the two multicenter studies, 5% of patients (19 of 369) discontinued modafinil due to an adverse event. The reasons for discontinuation included (in order): headache, cataplexy, nausea, depression, and nervousness.

3.1.4.4.4 Modafinil Status

Modafinil, in single or repeated doses ranging from 200 to 300 mg, improves cognitive performance and alertness during sleep deprivation. However, based on studies to date, it is not clear whether modafinil (at these doses) restores performance and alertness to non-sleep-deprived levels, or whether, based on a direct comparison, modafinil would be more effective than high-dose caffeine in the field. Tantalizing findings suggest some uniquely positive effects of modafinil, especially (a) its putative ability to actually reduce the need for recovery sleep, and (b) its apparent lack of disruptive effects on recovery sleep or naps. Replication and further study are needed to rule out alternative interpretations. Also, further studies on the effects of modafinil on self-assessment and judgment are warranted.

3.1.4.5 Sleep Inducers

The restorative effects of sleep are well known and well documented (Horne, 1988). Although the physiological basis of recuperation during sleep (here defined as the reestablishment of rested, pre-sleep deprivation performance levels) is unknown, the polysomnographic and behavioral parameters that reflect recuperation during sleep have been determined, with the most important being "sleep duration" (Wesensten et al., 1999). Simply stated, normal daytime alertness and performance levels are maintained by adequate nightly sleep.

Typically, both military and civilian operational demands preclude adequate sleep. Although, as indicated in the preceding section, stimulants can be used to temporarily boost performance when sleep is not possible, true and full restoration of performance and alertness occurs only with adequate recovery sleep. Ultimately, operator effectiveness during continuous operations (e.g., over weeks or months) depends on the adequacy of the sleep obtained during the operation.

One of the reasons that sleep tends to be inadequate during continuous operations is that, even though they are sleep deprived, operators are not always able to take full advantage of emergent opportunities for sleep. Several factors can interfere with adequate recovery sleep (e.g., if sleep initiation is attempted during the ascending phase of the circadian alertness rhythm, and/or if the environment is not conducive to sleep because of noise, light, motion, etc.). It is under these operational circumstances that pharmacologic enhancement of sleep may be desirable.

Among the currently available pharmacologic sleep inducers, benzodiazepine (BZ) agonists are currently the most widely prescribed because of their proven efficacy and relative safety (compared, for example, to older, non-BZ agonist sleep inducers such as barbiturates). Synthetic BZ agonists bind with high specificity to a homogeneous class of receptors in the brain called BZ receptors (Mohler & Okada, 1977; Squires & Braestrup, 1977). Many studies have demonstrated the efficacy with which BZ agonists hasten sleep onset and increase sleep duration. Examples of BZ agonists include triazolam (Halcion®), zolpidem (Ambien®[USA], Stilnox® [Europe], which, technically, is not in the benzodiazepine class of drugs but nevertheless acts as an agonist at the BZ receptor), and temazepam (Restoril® [USA], Normison® [Europe]).

Because the sleep and performance effects of most benzodiazepine agonists are qualitatively similar (with differences attributable primarily to differential pharmacokinetic profiles), a focus of this section will be on zolpidem – currently the most widely prescribed sleep-inducing medication. Some discussion will also be devoted to melatonin, a hormone secreted by the pineal gland during darkness and currently receiving significant attention as a "natural" sleep inducer, although the mechanism of action is not fully understood.

3.1.4.6 Zolpidem

Zolpidem (SL 80.0750-23N, N,N,6-trimethyl-2-(4-methyl-phenyl)imidazo[1,2-a]pyridine-3-acetamide hemitartrate) is manufactured in Europe under the trade name Stilnox® and in the USA under the trade name Ambien®. Zolpidem selectively binds to the central benzodiazepine-1 (BZ1) receptor (Langer & Arbilla, 1988). It is highly bound to plasma proteins and is transformed into inactive metabolites primarily by oxidation of methyl groups to carboxylic acids (Bianchetti et al., 1988; Thénot et al., 1988).

3.1.4.6.1 Hypnotic Efficacy of Zolpidem

Several laboratory studies have demonstrated that latency to sleep onset is significantly shortened by doses of zolpidem (ranging from 5 to 30 mg) compared to placebo during night-time administration (Lund, Ruther, Wober, & Hippius, 1988; Merlotti et al., 1988), following administration under non-sleep-conducive conditions (Balkin, O'Donnell, Wesensten, McCann, & Belenky, 1992; Walsh, Schweitzer, Muelbach, & Sugerman, 1988), and under conditions that mimic transient insomnia (Koshorek et al., 1988). Total sleep time (TST) has been shown to be increased by zolpidem at doses of 5, 10, 15, and 20 mg compared to placebo (Walsh et al., 1988), at doses of 7.5, 10.0 and 20.0 mg (Merlotti et al., 1988), and at a 30 mg dose (Nicholson & Pascoe, 1986, 1988). Zolpidem reduces the number of awakenings during sleep at doses of 20 and 30 mg (Nicholson & Pascoe, 1986, 1988) and at doses as low as 7.5 mg (Koshorek et al., 1988). Sleep efficiency (total sleep time/time in bed) has been found to increase across a range of zolpidem doses (5-20 mg, Koshorek et al., 1988; Vogel, Thurmond, MacIntosh, & Clifton, 1988); 20 and 30 mg, Nicholson & Pascoe, 1986, 1988). Sleep architecture has been shown to be affected by various doses of zolpidem, with, for example, Stage 1 (non-restorative) sleep reduced by 15 mg zolpidem in middle aged subjects (mean = 48.1 years) during night-time administration. Stage 2 sleep was decreased with 20 and 30 mg zolpidem in young subjects (mean = 20.9 years) during night-time administration. However, a 30 mg dose of zolpidem increased Stage 2 sleep in middle-aged subjects during night-time administration. The amounts of Stage 3 and Stage 4 sleep (i.e., deep sleep) have been found to be increased by 20 and 30 mg zolpidem (Koshorek et al., 1988; Nicholson & Pascoe, 1986, 1988), whereas the REM sleep amount was reduced by zolpidem at doses of 15 mg (Koshorek et al., 1988), 20 mg (Koshorek et al., 1988; Merlotti et al., 1988; Nicholson & Pascoe, 1986, 1988), and 30 mg (Nicholson & Pascoe, 1986, 1988).

3.1.4.6.2 Residual Effects of Zolpidem and Other Benzodiazepine Agonists on Performance

Results generally indicate that the short-acting sleep-inducing drugs such as triazolam (0.5 mg) and zolpidem (20 mg) substantially improve sleep under simulated operational conditions (Balkin et al., 1992). Although in most studies it has been reported that night-time administration of zolpidem does not impair next-day cognitive performance or alertness (perhaps due to its short, 1.5-2.4 h half-life; Fairweather, Kerr, & Hindmarch, 1992; Sicard, Trocherie, Moreau, Vieillefond, & Court, 1993), the few studies that have examined the performance effects of these drugs at or near the time of peak blood concentrations reveal that performance impairment is significant, and is positively correlated with sleep induction efficacy. Impairments have been shown in a variety of mental abilities including memory (Balkin, O'Donnell, Wesensten, & Belenky, 1991; Berlin et al., 1993; Wesensten, Balkin, & Belenky, 1995) and psychomotor functioning (Berlin et al., 1993).

Therefore, the likelihood that an operator might be required to awaken unexpectedly and quickly perform challenging tasks with a high degree of proficiency should be considered before sleep-inducing medications are administered in the operational environment. If this possibility exists, then it would be advisable to have the BZ antagonist, flumazenil, on hand for emergencies in which rapid reversal of BZ-agonist-induced decrements in alertness and performance is needed (Wesensten, Balkin, Davis, & Belenky, 1995).

3.1.4.6.3 Mechanism of Action, Pharmacokinetics, and Side Effects

It is likely that the pronounced hypnoselective profile of zolpidem arises from its action as a full agonist at the $GABA_A$ receptor subtype exhibiting selective BZ_1 (Benzodiazepine subtype 1) receptor binding. The absorption of zolpidem in humans appears to be complete with an absolute bioavailability of about 70% (Table 9).

Table 9: Pharmacokinetic Profile of Zolpidem (10 mg), according to Fraisse, Garrigou-Gadenne, and Thénot (1996)

Elimination half-life (h)	1.7 ± 0.1
C_{max} (µg/L)	139 ± 11.7
Bioavailability F	66.6 ± 4.4
t_{max} (h)	1.03 ± 0.02
AUC $_{0\text{-}12h}$ (µg/L/h)	483 ± 57

C_{max} values of about 120 µg.L^{-1}(or ng/ml) are usually attained between 30 and 60 minutes, which results in rapid sleep onset. For this reason, zolpidem should be administered just before bedtime. Zolpidem distributes homogeneously in the various tissues of the organism. Zolpidem and its metabolites are quickly eliminated from these tissues, and by 3 h after dosing, only residual amounts may be observed and only in the excretory organs. Zolpidem very rapidly crosses the blood-brain barrier with first-pass penetration into the brain that is quite high (a brain uptake index of 67% in the rat). Afterward, the efflux of zolpidem from the CNS is also very rapid. Zolpidem is extensively metabolized (oxidized) and the metabolites identified do not possess any pharmacological activity. They are eliminated in the urine.

Zolpidem does not impair cognitive performance and short-term memory on the morning after administration of a 10 or 20 mg dose at bedtime. It has been reported to produce some side effects such as vertigo, feelings of "empty headedness", drowsiness, headaches, and gastro-intestinal symptoms (Holm & Goa, 2000). In the case of chronic use by insomniac patients, no rebound of insomnia has been reported after withdrawal (Darcourt, Pringuey, Sallière, & Lavoisy, 1999). Neither pharmacological tolerance nor withdrawal symptoms have been shown with zolpidem (Sauvanet et al., 1988).

3.1.4.6.4 Zolpidem Status

Zolpidem is a safe and effective hypnotic (sleep-inducing) drug, and an excellent candidate for use as a sleep inducer in the operational environment, including at high altitude where zolpidem improves sleep quality without adversely affecting respiration (Beaumont et al., 1996). It has been shown to improve night-time sleep with no evidence of significant drug hangover effects on the following morning. However, like all benzodiazepine (BZ) agonists (and probably like all sleep inducers regardless of mechanism), zolpidem significantly impairs psychomotor performance and memory at or near peak blood concentration levels. Therefore, use of zolpidem in the operational environment might best be limited to those times when it is known that the operator will be afforded several hours of rest with a very low probability of being unexpectedly called to duty. In addition, further work is suggested to develop an oral formulation of flumazenil – a BZ antagonist – to reverse the performance-impairing effects of zolpidem when needed (i.e., in emergency situations).

3.1.4.7 Melatonin

The pineal hormone melatonin has received much attention in both the scientific and popular press. Claims have been made that melatonin improves sleep and readjusts the circadian rhythms of some variables such as body temperature and sleep. However, the effect (and effectiveness) of melatonin may depend upon the interaction of several factors including (a) circadian phase, (b) timing of ambient light exposure, and (c) extant sleep debt. For example, it is well established that time of administration impacts the effectiveness of melatonin. It must be taken at the correct time of day to resynchronize body temperature in the desired direction (advance or delay). Generally, the objective effects on sleep (latency and duration) are weak compared to prescription sleep-inducing agents such as zolpidem (Ambien®), triazolam (Halcion®), and temazepam (Restoril®). Also, there is no objective evidence that melatonin improves next-day cognitive performance, although it appears to improve subjective estimates of performance and well-being. More research is needed to determine melatonin's long-term effects and minimum effective dosage.

Melatonin appears to serve as a central nervous system (CNS) marker of day length. Results from most melatonin studies conducted over the past several decades generally indicate that exogenous melatonin produces some sedation, fatigue, and decreased alertness (subjective effects), it impairs response speed, and it shortens latency to sleep (objective effects). However, the extent to which melatonin increases sleep duration is still unclear.

3.1.4.7.1 Hypnotic Efficacy of Melatonin

Study of the putative sedative effect of endogenously administered (intravenous) melatonin in humans dates back to Aaron Lerner's laboratory in 1960 after he identified and named melatonin (Lerner & Case, 1960). Additional studies soon were undertaken following several indirect signs linking sleep and melatonin production (e.g., animal studies that showed night-time elevation of pineal enzymes that synthesize melatonin and the assay methods that allowed the measurement of melatonin concentrations and its metabolite to assess pineal function in humans).

Many studies have documented the sleep-promoting and sleep-inducing effects of pharmacological doses (i.e., higher than normal physiological levels) of melatonin in humans using a wide array of measurements (e.g., subjective self-reports, polysomnographic recordings, actigraphic recordings of motor activity, multiple sleep latency tests, and psychological and performance testing). With few exceptions, the results from each of these studies suggest that a substantial increase in circulating melatonin levels is associated with sedation, fatigue, decreased alertness, significantly increased reaction time, a shortened sleep onset latency, increased sleep efficiency and total sleep time, and/or increased sleep propensity. These effects appear to be specific to the doses utilized and the time of day administered. Originally, it was thought that high doses (10, 20, 40, 80, and 240 mg) were necessary to induce sleepiness and sleep during the daytime (Dollins et al., 1993; Lieberman, Waldhauser, Garfield, Lynch, & Wurtman, 1984) and late in the evening (Waldhauser, Saletu, & Trinchard-Lugan, 1990). However, lower pharmacological doses (1-6 mg) have been found to produce sleep-inducing effects in the afternoon (Dijk et al., 1995; Dollins, Zhdanova, Wurtman, Lynch, & Deng, 1994; Rogers, Phan, Kennaway, & Dawson, 1998) or, in some cases, increased evening fatigue (Nave, Peled, & Peretz, 1995; Zhdanova et al., 1995; Zhdanova, Wurtman, Morabito, Piotrovska, & Lynch, 1996). In some studies, melatonin failed to affect the onset or duration of sleep (James, Mendelson, Sack, Rosenthal, & Wehr, 1987; Mishima, Satoh, Shimizu, & Hishikawa, 1997), although a "ceiling effect" (i.e., an inability to improve upon "normal, good" sleep) is possible. In general, regardless of the dose, melatonin's sleep-promoting effect typically is manifested within 30 to 60 min after administration.

Initially, mostly pharmacological doses were tested. When circulating concentrations of melatonin were measured, they were found to range from several-fold to several thousand-fold above the levels that occur naturally in humans. When melatonin doses less than 1 mg were tested, a dose dependency of the sleep-promoting effect was revealed (Dollins et al., 1994). Melatonin doses of 0.1, 0.3, 1.0, and 10.0 mg have been tested, and shown to increase subjective sleepiness or shortened latency to sleep onset, although the 0.1 mg dose was less potent than the 0.3 mg or higher doses. However, even the two lower doses were sufficient to increase circulating melatonin (87.7 and 213.2 pg/ml, respectively) to levels within the normal nocturnal physiological range (0-200 pg/ml) in adult humans, and to promote sleepiness and sleep.

In studies by Zhdanova and colleagues, the effects of 0.3 to 1.0 mg doses of melatonin and placebo were compared, and results confirmed that increasing the circulating melatonin levels to within the physiological range promotes polysomnographically measured sleep onset in both afternoon naps (Zhdanova et al., 1995) and overnight sleep (Zhdanova et al., 1996) in young, healthy volunteers.

Independent of the dose used, the sleep-promoting effect of melatonin is not characterized by changes in the duration or relative amounts of the various sleep stages (Anton-Tay, 1974; Waldhauser et al., 1990). However, Attenburrow, Cowen, & Sharpley (1996) more recently reported that a 1.0 mg dose of melatonin (but not 0.3 mg) significantly increased total sleep time and sleep efficiency in healthy, middle-aged subjects. This report suggests a supra-physiological threshold (i.e., higher than normal physiological levels) for sleep-promoting effects.

The soporific effects of melatonin following daytime administration have been documented with objective EEG monitoring as well as with subjective measures. Tzischinsky and Lavie (1984) administered 5 mg melatonin at varying times (1200, 1700, 1900 and 2100) across separate days and found that melatonin significantly increased sleep propensity, the spectral power in the theta, delta, and spindle bands, and subjective sleepiness. The latency to maximum effect varied from 1 hour at 2100 hrs to 3 hours and 4 minutes at 1200 hrs. These findings were replicated by Nave, Herer, Haimov, Shlitner, and Lavie (1996). In their study, 3 mg melatonin administered at 1200 hrs significantly decreased the latency to fall asleep and increased total sleep time until 1900 hrs. Similar findings were reported by Reid, Van Den Heuvel, & Dawson (1996) using a modified version of the Multiple Sleep Latency Test (MSLT) after administration of 5 mg melatonin at 1400 hrs. When administered at 1300 or 1800 hrs, 5 mg melatonin significantly increased subjective sleepiness and modified waking EEG power density in the theta/alpha range (Cajochen et al., 1996). However, the subjective effects appeared 40-90 min after administration, whereas the effects on EEG appeared almost immediately. In two studies, the effects of melatonin on daytime naps (two-hour and four-hour) were examined, and it was found that doses ranging from 1 to 10 mg improved sleep efficiency and decreased sleep latency (Hughes & Badia, 1997; Nave et al., 1995). There also is evidence that melatonin modifies sleep EEG activity during the day in a manner that is similar to the benzodiazepines, via enhanced EEG power density in the sigma range during non-REM sleep (Dijk et al., 1995).

3.1.4.7.2 Melatonin and the Circadian Timing System

An interesting and plausible hypothesis that accounts for variability in the effectiveness of melatonin associated with time of day has been proposed by Sack, Hughes, Edgar, and Lewy (1997). They suggest that the soporific effects of melatonin are the result of its actions on the suprachiasmatic nucleus (SCN). These actions have been identified as phase shifting of the circadian pacemaker, located in the SCN, and/or attenuation/antagonism of the SCN-dependent mechanism that promotes and maintains cortical and behavioral activation at particular times of day. Both of these possible effects are presumed to occur at physiological doses of melatonin. If only higher, pharmacological doses of melatonin are effective for sleep promotion, then it would be unlikely that endogenous melatonin played a role in normal sleep processes (Sack et al., 1997).

3.1.4.7.3 Melatonin Effects on Performance

In most studies to determine the cognitive performance effects of melatonin, global subjective measures have been utilized (Nickelsen, Demisch, Radermacher, & Schoffling, 1989; Nickelsen, Lang, & Bergau, 1991; Zhdanova et al., 1995). Inferential measures have also been used, including a small range of tests (Arendt, Borbely, & Wright, 1984; Wynn & Arendt, 1988), often involving relatively limited subject numbers (Wynn & Arendt, 1988; Zhdanova et al., 1995). To date, some of the very few studies to report performance effects indicate a significant decrease in performance (i.e., an increase in mean response time scores) on visual choice performance tasks (Dollins et al., 1993, 1994; Lieberman, Wurtman, & Teicher, 1989; Rogers et al., 1998) with melatonin doses ranging from 1 to 80 mg. Mean reaction time scores for a visual choice task were also increased in the Rogers et al. study but not in the Lieberman et al. and Dollins et al. studies, possibly because the former study utilized a two-choice task while the latter studies utilized a four-choice task.

3.1.4.7.4 Mechanism of Action, Pharmacokinetics, and Side Effects

Melatonin (N-acetyl-5-methoxytryptamine) is normally secreted during the dark phase of the day in all species studied. It is synthesized from serotonin through two enzymatic steps: first is the N-acetylation by serotonin N-acetyltransferase (SNAT) to yield N-acetylserotonin; second is the transfer of a methyl group from S-adenosylmethionine to the 5-hydroxy group of N-acetylserotonin to yield melatonin. It appears that melatonin is predominantly secreted by the pineal gland by simple diffusion and that the concentration of melatonin in the pineal is a direct reflection of its synthesis and its concentration in plasma (Redman, 1997).

Depending on the species, intravenously administered radioactive melatonin rapidly disappears from the blood with a half-life of about 30 min (Pang, Lee, Chan, & Ayre, 1993). There is considerable variability in the half-life for humans, with a range between 35 and 50 min (Vakkuri, Leppaluoto, & Kauppila, 1985; Waldhauser et al., 1984). Time to peak plasma concentration is typically 60 min and single bolus formulations produce physiological levels in the blood which are maintained for 2 to 4 hours.

Melatonin doses of 1 to 5 mg usually produce high physiological night-time plasma levels for 3-8 hours. Oral doses of melatonin up to 0.5 mg produce plasma levels that approximate the range of endogenous melatonin (0-200 pg/ml). Not surprisingly, slow-release and fast-release formulations result in varied absorption and blood levels. A dose of 0.5 mg melatonin is generally considered to mark the cut-off between "high physiological" and "low pharmacological" doses, although in many subjects, even a 0.5 mg dose can elevate melatonin blood levels into the pharmacological range (i.e., higher than normal, physiological levels – also referred to as the "supra-physiological" range).

3.1.4.7.5 Melatonin Status

A considerable number of studies have been conducted to determine the efficacy of melatonin for induction of sleep and/or facilitation of adaptation to new time zones. These studies have generally suggested that the effects of melatonin are both dose-dependent and "time of day"-dependent. Direct comparisons with BZ agonists are lacking, but based on the magnitude of reported effects, it appears that both the sleep-inducing and performance-impairing effects of melatonin are comparatively mild.

3.1.4.8 Summary

Since, for the operator, sleep loss is probably the most common and salient consequence of modern military and civilian operations, pharmacological agents that optimize control over alertness and sleep will constitute vital components of any armamentarium assembled for the purpose of optimizing operator functional capacity. Stimulants are most useful during short-term sustained operations, when performance must be maintained without benefit of sleep. Of the currently available stimulants, caffeine and modafinil are the most promising because of their efficacy, safety, and low abuse potential. Sleep inducers are most useful during continuous operations (lasting weeks or months), when there is opportunity to sleep, but sleep is inadequate due to circadian desynchrony, non-sleep-conducive environmental factors, etc. Of the currently available hypnotics (or putative hypnotics), BZ agonists such as zolpidem, and to a lesser extent the hormone melatonin, are the most promising, again because of their relative safety, efficacy, and low abuse potential. However, further research is needed to determine the schedules, doses, and combinations that will allow utilization of these pharmacological agents to maximum benefit in the operational environment.

3.1.4.9 References

Akerstedt, T., & Ficca, G. (1997). Alertness-Enhancing Drugs as a Countermeasure to Fatigue in Irregular Work Hours. *Chronobiology International*, *14*, 145-158.

Anton-Tay, F. (1974). Melatonin: Effects on Brain Function. In E. Costa, G.L. Gessa, & M. Sandler (Eds.), *Serotonin – New Vistas: Biochemistry and Behavioral and Clinical Studies (Advances in Biochemical Pharmacology)* (Vol. 11). New York: Raven Press.

Arendt, J., Borbely, A.A., & Wright, J. (1984). The Effects of Chronic, Small Doses of Melatonin given in the Late Afternoon on Fatigue in Man: A Preliminary Study. *Neuroscience Letters*, *45*, 317-321.

Attenburrow, M.E.J., Cowen, P.J., & Sharpley, A.L. (1996). Low Dose Melatonin Improves Sleep in Healthy Middle-Aged Subjects. *Psychopharmacology*, *126*, 179-181.

Babkoff, H., Mikulincer, M., Caspy, T., Carasso, R.L., & Sing, H.C. (1989). The Implications of Sleep Loss for Circadian Performance Accuracy. *Work Stress*, *3*, 3-14.

Babkoff, H., Sing, H.C., Thorne, D.R., Genser, S.G., & Hegge, F.W. (1989). Perceptual Distortions and Hallucinations Reported during the Course of Sleep Deprivation. *Perceptual Motor Skills*, *68*, 787-798.

Balkin, T.J., O'Donnell, V.M., Wesensten, N.J., & Belenky, G. (1991). Somnogenically Equivalent Doses of Zolpidem and Triazolam I: Effects on Restricted Reminding Memory Test Performance. *Sleep Research*, *20*, 58.

Balkin, T.J., O'Donnell, V.M., Wesensten, N., McCann, U., & Belenky, G. (1992). Comparison of the Daytime Sleep and Performance Effects of Zolpidem vs. Triazolam. *Psychopharmacology*, *107*, 83-88.

Baranski, J.V., & Pigeau, R.A. (1997). Self-Monitoring Cognitive Performance during Sleep Deprivation: Effects of Modafinil, *d*-amphetamine and Placebo. *Journal of Sleep Research*, *6*, 84-91.

Batéjat, D., & Lagarde, D. (1999). Naps and Modafinil as Countermeasures for the Effects of Sleep Deprivation on Cognitive Performance. *Aviation, Space, and Environmental Medicine*, *70*, 493-498.

Beaumont, M., Batéjat, D., Piérard, C., Coste, O., Doireau, P., Van Beers, P., Chauffard, F., Chassard, D., Enslen, M., Denis, J.B., & Lagarde, D. (2001). Slow Release Caffeine and Prolonged (64-h) Continuous Wakefulness: Effects on Vigilance and Cognitive Performance. *Journal of Sleep Research*, *10*, 265-276.

Beaumont, M., Goldenberg, F., Lejeune, D., Marotte, H., Harf, A., & Lofaso, F. (1996). Effect of Zolpidem on Sleep and Ventilatory Patterns at Simulated Altitude of 4,000 Meters. *American Journal of Respiratory and Critical Care Medicine*, *153*, 1864-1869.

Béracochéa, D., Cagnard, B., Célérier, A., Le Merrer, J., Pérès, M., & Piérard, C. (2001). First Evidence of a Delay-Dependent Working Memory-Enhancing Effect of Modafinil in Mice. *Neuroreport*, *12*, 375-378.

Berlin, I., Warot, D., Hergueta, T., Molinier, P., Bagot, C., & Pucch, A.J. (1993). Comparison of the Effects of Zolpidem and Triazolam on Memory Functions, Psychomotor Performances, and Postural Sway in Healthy Subjects. *Journal of Clinical Psychopharmacology*, *13*, 100-106.

Bianchetti, G., Dubruc, C., Thiercelin, J.F., Bercoff, E., Bouchet, E.L., Emeriau, J.P., Galperine, I., Lambert, D., Vandel, B., & Thébault, J.J. (1988). Clinical Pharmacokinetics of Zolpidem in various Physiological and Pathological Conditions. In G. Bartholini, J.C. Friedmann, S.Z. Sanger, P.L. Morselli, & A. Wick (Eds.), *Imidazopyridines in Sleep Disorders: A Novel Experimental and Therapeutic Approach* (pp. 155-164). New York: Raven Press.

Bonati, M., Latini, R., Galletti, F., Young, J.F., Tognoni, G., & Garattini, S. (1982). Caffeine Disposition after Oral Doses. *Clinical and Pharmacological Therapeutics*, *32*, 98-106.

Bonnet, M.H., & Arand, D.L. (1994). Impacts of Naps and Caffeine on Extended Nocturnal Performance. *Physiology and Behavior*, *56*, 103-109.

Brun, J., Chamba, G., Khalfallah, Y., Girard, P., Boissy, I., Bastuji, H., Sassolas, G., & Claustrat, B. (1998). Effect of Modafinil on Plasma Melatonin, Cortisol and Growth Hormone Rhythms, Rectal Temperature and Performance in Healthy Subjects during a 36 h Sleep Deprivation. *Journal of Sleep Research*, *7*, 105-114.

Buguet, A., Montmayeur, A., Pigeau, R., & Naitoh, P. (1995). Modafinil, *d*-amphetamine and Placebo during 64 hours of Sustained Mental Work. II. Effects on Two Nights of Recovery Sleep. *Journal of Sleep Research*, *4*, 229-241.

Cajochen, C.H., Krauchi, K., von Arx, M., Mori, D., Graw, P., & Wirtz-Justice, A. (1996). Daytime Melatonin Administration Enhances Sleepiness and Theta/Alpha Activity in the Waking EEG. *Neuroscience Letters*, *207*, 209-213.

Caldwell, J.A., Caldwell, J.L., Smythe, N.K., & Hall, K.K. (2000). A Double-Blind, Placebo-Controlled Investigation of the Efficacy of Modafinil for Sustaining the Alertness and Performance of Aviators: A Helicopter Simulator Study. *Psychopharmacology*, *150*, 272-282.

Caldwell, J.A., Smythe, N.K., Leduc, P.A., & Caldwell, J.L. (2000). Efficacy of Dexedrine (R) for Maintaining Aviator Performance during 64 hours of Sustained Wakefulness: A Simulator Study. *Aviation, Space, and Environmental Medicine*, *71*, 7-18.

Darcourt, G., Pringuey, D., Sallière, D., & Lavoisy, J. (1999). The Safety and Tolerability of Zolpidem – An Update. *Journal of Psychopharmacology*, *13*, 81-93.

Dews, P.B. (1984a). Metabolism and Kinetics. In P.B. Dews (Ed.), *Caffeine* (pp. 19-75). New York: Springer-Verlag.

Dews, P.B. (1984b). Behavioral Effects of Caffeine. In P.B. Dews (Ed.), *Caffeine* (pp. 86-103). New York: Springer-Verlag.

Dijk, D.J., Roth, C., Landolt, H.P., Werth, E., Aeppli, M., Achermann, P., & Borbely, A.A. (1995). Melatonin Effect on Daytime Sleep in Men: Suppression of EEG Low Frequency Activity and Enhancement of Spindle Frequency Activity. *Neuroscience Letters*, *201*, 13-16.

Dollins, A.B., Lynch, H.J., Wurtman, R.J., Deng, M.H., Kischka, K.U., Gleason, R.E., & Lieberman, H.R. (1993). Effect of Pharmacological Daytime Doses of Melatonin on Human Mood and Performance. *Psychopharmacology*, *112*, 490-496.

Dollins, A.B., Zhdanova, I.V., Wurtman, R.J., Lynch, H.J., & Deng, M.H. (1994). Effect of Inducing Nocturnal Serum Melatonin Concentrations in Daytime on Sleep, Mood, Body Temperature, and Performance. *Proceedings of the National Academy of Science USA*, *91*, 1824-1828.

Fairweather, D.B., Kerr, J.S., & Hindmarch, I. (1992). The Effects of Acute and Repeated Doses of Zolpidem on Subjective Sleep, Psychomotor Performance and Cognitive Function in Elderly Volunteers. *European Journal of Clinical Pharmacology*, *43*, 597-601.

Fraisse, J., Garrigou-Gadenne, D., & Thénot, J.P. (1996). Pharmacokinetic and Metabolic Profiles of Zolpidem. In H. Freeman, A.J. Puech, & T. Roth (Eds.), *Zolpidem: An Update of its Pharmacological Properties and Therapeutic Place in the Management of Insomnia* (pp. 45-57). Paris: Elsevier.

Holm, K.J., & Goa, K.L. (2000). Zolpidem. An Update of its Pharmacology, Therapeutic Efficacy and Tolerability in the Treatment of Insomnia. *Drugs*, *59*, 865-889.

Horne, J. (1988). *Why We Sleep. The Functions of Sleep in Humans and Other Mammals*. Oxford: Oxford University Press.

Hughes, R.J., & Badia, P. (1997). Sleep-Promoting and Hyperthermic Effects of Daytime Melatonin Administration in Humans. *Sleep*, *20*, 124-131.

Institute of Medicine. (2001). *Caffeine for the Sustainment of Mental Performance: Formulations for Military Operations*. Washington, DC: National Academy Press.

James, S.P., Mendelson, W.B., Sack, D.A., Rosenthal, N.E., & Wehr, T.A. (1987). The Effect of Melatonin on Normal Sleep. *Neuropsychopharmacology*, *1*, 41-44.

Kamimori, G.H., Eddington, N.A., Hoyt, R.W., Fulco, C.S., Lugo, S., Durkot, M.J., Brunhart, A.E., & Cymerman, A. (1995). Effects of Altitude (4300 m) on the Pharmacokinetics of Caffeine and Cardio-Green in Humans. *European Journal of Clinical Pharmacology*, *48*, 167-170.

Kamimori, G.H., Karyekar, C., Otterstetter, R., Cox, D.S., Balkin, T.J., Belenky, G.L., & Eddington, N.D. (2002). The Rate of Absorption and Relative Bioavailability of Caffeine administered in Chewing Gum versus Capsules to Normal Healthy Volunteers. *International Journal of Pharmaceutics*, *234*, 159-167.

Kamimori, G.H., Lugo, S.I., Penetar, D.M., Chamberlain, A.C., Brunhart, G.E., Brunhart, A.E., & Eddington, N.D. (1995). Dose-Dependent Caffeine Pharmacokinetics during Severe Sleep Deprivation in Humans. *International Journal of Clinical Pharmacology, Therapy, and Toxicology*, *33*, 182-186.

Kaplan, G.B., Greenblatt, D.J., Ehrenberg, B.L., Goddard, J.E., Cotreau, M.M., Harmatz, J.S., & Shader, R.I. (1997). Dose-Dependent Pharmacokinetics and Psychomotor Effects of Caffeine in Humans. *Journal of Clinical Pharmacology*, *37*, 693-703.

Koshorek, G., Roehrs, T., Sicklesteel, J., Merlotti, L., Russo, L., & Roth, T. (1988). Dose Effects of Zolpidem on Transient Insomnia. *Sleep Research*, *17*, 47.

Lagarde, D., Batéjat, D., Sicard, B., Trocherie, S., Chassard, D., Enslen, M., & Chauffard, F. (2000). Slow Release Caffeine: A New Response to the Effects of a Limited Sleep Deprivation. *Sleep*, *23*, 651-661.

Lagarde, D., Batéjat, D., Van Beers, P., Sarafian, D., & Pradella, S. (1995). Interest of Modafinil, a new Psychostimulant, during a Sixty-Hour Sleep Deprivation Experiment. *Fundamental and Clinical Pharmacology*, *9*, 271-279.

Langer, S.Z., & Arbilla, S. (1988). Limitations of the Benzodiazepine Receptor Nomenclature: A Proposal for a Pharmacological Classification as Omega Receptor Subtypes. *Fundamental and Clinical Pharmacology*, *2*, 159-170.

Lerner, A.B., & Case, J.D. (1960). Melatonin. *Federal Proceedings*, *19*, 590-592.

Lieberman, H., Waldhauser, F., Garfield, G., Lynch, H., & Wurtman, R. (1984). Effects of Melatonin on Human Mood and Performance. *Brain Research*, *323*, 201-207.

Lieberman, H.R., Wurtman, J.J., & Teicher, M.H. (1989). Circadian Rhythms of Activity in Healthy Young and Elderly Humans. *Neurobiology of Aging*, *10*, 259-265.

Lorist, M.M., & Snel, J. (1997). Caffeine Effects on Perceptual and Motor Processes. *Electroencephalography and Clinical Neurophysiology*, *102*, 401-413.

Lund, R., Ruther, E., Wober, W., & Hippius, H. (1988). Effects of Zolpidem (10 and 20 mg), Lormetazepam, Triazolam, and Placebo on Night Sleep and Residual Effects during the Day. In G. Bartholini, J.C. Friedmann, S.Z. Sanger, P.L. Morselli, & A. Wick (Eds.), *Imidazopyridines in Sleep Disorders: A Novel Experimental and Therapeutic Approach* (pp. 193-204). New York: Raven-Press.

Lyons, T.J., & French, J. (1991). Modafinil: The Unique Properties of a new Psychostimulant. *Aviation, Space, and Environmental Medicine*, *62*, 432-435.

Merlotti, L., Roehrs, T., Koshorek, G., Zorick, F., Lamphere, J., & Roth, T. (1988). The Dose Effects of Zolpidem on the Sleep of Healthy Normals. *Sleep Research*, *17*, 51.

Mishima, K., Satoh, K., Shimizu, T., & Hishikawa, Y. (1997). Hypnotic and Hypothermic Action of Daytime-Administered Melatonin. *Psychopharmacology*, *133*, 168-171.

Moachon, G., Kanmacher, I., Clenet, M., & Matinier, D. (1996). Pharmacokinetic Profile of Modafinil. *Drugs of Today*, *32(4)*, 32-37.

Mohler, H., & Okada, T. (1977). Benzodiazepine Receptor: Demonstration in the Central Nervous System. *Science*, *198*, 849-851.

Nave, R., Herer, P., Haimov, I., Shlitner, A., & Lavie, P. (1996). Hypnotic and Hypothermic Effects of Melatonin on Daytime Sleep in Humans: Lack of Antagonism by Flumazenil. *Neuroscience Letters*, *214*, 123-126.

Nave, R., Peled, R., & Peretz, L. (1995). Melatonin Improves Evening Napping. *European Journal of Pharmacology*, *275*, 213-216.

Newcombe, P.F., Renton, K.W., Rautaharju, P.M., Spencer, C.A., & Montague, T.J. (1988). High-Dose Caffeine and Cardiac Rate and Rhythm in Normal Subjects. *Chest*, *94*, 90-94.

Nicholson, A.N., & Pascoe, P.A. (1986). Hypnotic Activity of an Imidazo-Pyridine (Zolpidem). *British Journal of Clinical Pharmacology*, *21*, 205-211.

Nicholson, A.N., & Pascoe, P.A. (1988). Hypnotic Activity of Zolpidem: Night-Time and Daytime Studies in Young and Middle-Aged Adults. In G. Bartholini, J.C. Friedmann, S.Z. Sanger, P.L. Morselli, & A. Wick (Eds.), *Imidazopyridines in Sleep Disorders: A Novel Experimental and Therapeutic Approach* (pp. 231-240). New York: Raven Press.

Nickelsen, T., Demisch, K., Radermacher, B., & Schoffling, K. (1989). Influence of Subchronic Intake of Melatonin at Various Times of the Day on Fatigue and Hormonal Levels: A Placebo-Controlled, Double-Blind Trial. *Journal of Pineal Research*, *6*, 325-334.

Nickelsen, T., Lang, A., & Bergau, L. (1991). The Effects of 6-, 9- and 11-hour Time Shifts on Circadian Rhythms: Adaptation of Sleep Parameters and Hormonal Patterns following the Intake of Melatonin or Placebo. *Advances in Pineal Research*, *5*, 303-306.

Pang, S.-F., Lee, P.N., Chan, Y.S., & Ayre, E.A. (1993). Melatonin Secretion and its Rhythms in Biological Fluids. In H.S. Yu & R.J. Reiter (Eds.), *Melatonin: Biosynthesis, Physiological Effects, and Clinical Applications* (pp. 129-153). Boca Raton, FL: CRC Press.

Patat, A., Rosenzweig, P., Enslen, M., Trocherie, S., Miget, N., Bozon, M.C., Allain, H., & Gandon, J.M. (2000). Effects of a new Slow Release Formulation of Caffeine on EEG, Psychomotor and Cognitive Functions in Sleep-Deprived Subjects. *Human Psychopharmacology Clinical and Experimental*, *15*, 153-170.

Penetar, D., McCann, U., Thorne, D., Kamimori, G., Galinski, G., Sing, H., Thomas, M., & Belenky, G. (1993). Caffeine Reversal of Sleep Deprivation Effects on Alertness and Mood. *Psychopharmacology*, *112*, 359-365.

Penetar, D., McCann, U., Thorne, D., Schelling, A., Galinski, C., Sing, H., Thomas, M., & Belenky, G. (1994). Effects of Caffeine on Cognitive Performance, Mood and Alertness in Sleep-Deprived Humans. In *Food Components to Enhance Performance*. Washington, DC: National Academic Press.

Piérard, C., Beaumont, M., Enslen, M., Chauffard, F., Tan, D.X., Reiter, R.J., Fontan, A., French, J., Coste, O., & Lagarde, D. (2001). Resynchronization of Hormonal Rhythms after an Eastbound Flight in Humans: Effects of Slow-Release Caffeine and Melatonin. *European Journal of Applied Physiology*, *85*, 144-150.

Pigeau, R., Naitoh, P., Buguet, A., McCann, C., Baranski, J., Taylor, M., Thompson, M., & Mack, I. (1995). Modafinil, *d*-amphetamine, and Placebo during 64 hours of Sustained Mental Work I: Effects on Mood, Fatigue, Cognitive Performance, and Body Temperature. *Journal of Sleep Research*, *4*, 212-228.

Pigeau, R.A. (2001). An Overview of Sleep Deprivation and the Ameliorative Effects of Modafinil. *Sleep/Wakefulness Management in Continuous/Sustained Operations* (Vol. RTO EN-016 / HFM-064, pp. 2.1-2.26). Neuilly-sur-Seine, France: NATO/RTO.

Redman, J.R. (1997). Circadian Entrainment and Phase Shifting in Mammals with Melatonin. *Journal of Biological Rhythms*, *12*, 581-587.

Reid, K., Van Den Heuvel, C., & Dawson, D. (1996). Day-Time Melatonin Administration: Effects on Core Temperature and Sleep Onset Latency. *Journal of Sleep Research*, *5*, 150-154.

Ritschel, W.A. (1986). *Handbook of Basic Pharmacokinetics* (3rd ed.). Hamilton, IL: Drug Intelligence Publications.

Robertson, D., & Curatolo, P.W. (1984). The Cardiovascular Effects of Caffeine. In P.B. Dews (Ed.), *Caffeine* (pp. 77-85). New York: Springer-Verlag.

Rogers, N.L., Phan, O., Kennaway, D.J., & Dawson, D. (1998). Effect of Daytime Oral Melatonin Administration on Neurobehavioral Performance in Humans. *Journal of Pineal Research*, *25*, 47-53.

Sack, R.L., Hughes, R.J., Edgar, D.M., & Lewy, A.J. (1997). Sleep-Promoting Effects of Melatonin: At What Dose, in Whom, under What Conditions, and by What Mechanisms? *Sleep*, *20*, 908-915.

Sauvanet, J.P., Maarek, L., Roger, M., Renaudin, J., Louvel, E., & Orofiamma, B. (1988). Open Long-Term Trials with Zolpidem in Insomnia. In J.P. Sauvanet, S.Z. Langer, & P.L. Morselli (Eds.), *Imidazopyridines in Sleep Disorders* (pp. 339-350). New York: Raven Press.

Serafin, W.E. (1996). Drugs used in the Treatment of Asthma: Methylxanthines. In J.G. Hardman & L.E. Limbird (Eds.), *The Pharmacological Basis of Therapeutics* (9th ed., pp. 672-682). New York: Macmillan.

Sicard, B.A., Trocherie, S., Moreau, J., Vieillefond, H., & Court, L.A. (1993). Effects of Zolpidem on Alertness and Psychomotor Abilities among Aviation Ground Personnel and Pilots. *Aviation, Space, and Environmental Medicine*, *64*, 371-375.

Smith, A.P. (1995). Caffeine, Caffeine Withdrawal and Psychomotor Performance: A Reply to James. *Neuropsychobiology*, *31*, 200-201.

Squires, R.F., & Braestrup, C. (1977). Benzodiazepine Receptors in Rat Brain. *Nature*, *266*, 732-734.

Thénot, J.P., Hermann, P., Durand, A., Burke, J.T., Allen, J., Garrigou, D., Vajta, S., Albin, H., Thébault, J.J., Olive, G., & Warrington, S.J. (1988). Pharmacokinetics and Metabolism of Zolpidem in various Animal Species and in Humans. In G. Bartholini, J.C. Friedmann, S.Z. Sanger, P.L. Morselli, & A. Wick (Eds.), *Imidazopyridines in Sleep Disorders: A Novel Experimental and Therapeutic Approach* (pp. 139-154). New York: Raven Press.

Tzischinsky, O., & Lavie, P. (1984). Melatonin Possesses Time-Dependent Hypnotic Effects. *Sleep*, *17*, 638-645.

Vakkuri, O., Leppaluoto, J., & Kauppila, A. (1985). Oral Administration and Distribution of Melatonin in Human Serum, Saliva and Urine. *Life Sciences*, *37*, 489-495.

Vogel, G.W., Thurmond, A., MacIntosh, M., & Clifton, T. (1988). The Effects of Zolpidem on Transient Insomnia. *Sleep Research*, *17*, 67.

Waldhauser, F., Saletu, B., & Trinchard-Lugan, I. (1990). Sleep Laboratory Investigations on Hypnotic Properties of Melatonin. *Psychopharmacology*, *100*, 222-226.

Waldhauser, F., Waldhauser, M., Lieberman, H.R., Deng, M.H., Lynch, H.J., & Wurtman, R.J. (1984). Bioavailability of Oral Melatonin in Humans. *Neuroendocrinology*, *39*, 307-313.

Walsh, J.K., Schweitzer, P.K., Muelbach, M.J., & Sugerman, J.L. (1988). Experimental Phase-Advance as a Model of Transient Insomnia and Effects of Zolpidem. *Sleep Research*, *17*, 267.

Welling, P.G. (1986). Drug Absorption, Distribution, Metabolism and Excretion. In P.G. Welling (Ed.), *Pharmacokinetics: Process and Mathematics* (pp. 35-119). Washington, DC: American Chemical Society.

Wesensten, N.J., Balkin, T.J., & Belenky, G.L. (1995). Effects of Daytime Administration of Zolpidem versus Triazolam on Memory. *European Journal of Clinical Pharmacology*, *48*, 115-122.

Wesensten, N.J., Balkin, T.J., & Belenky, G. (1999). Does Sleep Fragmentation Impact Recuperation? A Review and Re-Analysis. *Journal of Sleep Research*, *8*, 237-245.

Wesensten, N.J., Balkin, T.J., Davis, H.Q., & Belenky, G.L. (1995). Reversal of Triazolam- and Zolpidem-Induced Memory Impairment by Flumazenil. *Psychopharmacology (Berl)*, *121*, 242-249.

Wisor, J.P., Nishino, S., Sora, I., Uhl, G.H., Mignot, E., & Edgar, D.M. (2001). Dopaminergic Role in Stimulant-Induced Wakefulness. *Journal of Neuroscience*, *21*, 1787-1794.

Wynn, V.T., & Arendt, J. (1988). Effect of Melatonin on the Human Electrocardiogram and Simple Reaction Time Responses. *Journal of Pineal Research*, *5*, 427-435.

Zhdanova, I.V., Wurtman, R.J., Lynch, H.J., Ives, J.R., Dollins, A.B., Morabito, C., Matheson, J.K., & Schomer, D.L. (1995). Sleep-Inducing Effects of Low Doses of Melatonin Ingested in the Evening. *Clinical Pharmacology and Therapeutics*, *57*, 552-558.

Zhdanova, I.V., Wurtman, R.J., Morabito, C., Piotrovska, V.R., & Lynch, H.J. (1996). Effects of Low Oral Doses of Melatonin, given 2-4 Hours before Habitual Bedtime, on Sleep in Normal Young Humans. *Sleep*, *19*, 423-431.

3.1.5 Sustained Acceleration

3.1.5.1 Definition and Measurement

Sustained acceleration, or G, refers to the exposure of pilots to greater (+Gz) or less (-Gz) than normal gravitational acceleration as a result of high speed maneuvering of aircraft (Burns, 1995a; Prior, 1995). With changes in Gz forces, the weight of the blood in various vessels of the body is increased or decreased (Burns, 1995b). With high +Gz, there is a drop in the hydrostatic pressure in those blood vessels above the heart (i.e., in the neck and facial blood vessels, and in the cerebral circulation). In addition, the increased weight of blood in the lower body will result in distension of the venous capacitance vessels and a decreased return of blood to the heart. As a result of the siphoning effect, the heart is not required to pump against the additional weight of the blood (Burton, 1965). However, the collapse of the jugular veins of the neck will result in a large increase in resistance to blood flow increases (non-linear with respect to the magnitude of +Gz), which appears to be the primary reason for G loss of consciousness (GLOC; Cirovic, Walsh, & Fraser, 2000; Cirovic, Walsh, & Fraser, 2001; Cirovic, Walsh, Fraser, & Gualino, 2002). The collapse of the cerebral veins is prevented by a corresponding decrease in cerebrospinal fluid pressure, thus preventing a fall in the transmural pressure (Rushmer, Beckman, & Lee, 1947).

In addition to the direct effects of sustained acceleration on the perfusion of the brain, pilots may experience severe limb pain (due to both Gz forces and anti-G protective equipment) as well as fatigue (due to repeated muscular straining against the Gz forces) and thermal and dehydration stress due to the additional layers of protective garments. Gz exposure and anti-Gz protective equipment can also impact pulmonary function and affect lung gas exchange (Prior, 1995).

3.1.5.2 Background

Research into the effects of acceleration on operator state was undertaken as far back as World War II (Wood, Lambert, Baldes, & Code, 1946), with the primary focus on the development of better techniques for the measurement of the efficacy of anti-G protective procedures (e.g., the anti-G straining maneuver) and life support equipment (e.g., the anti-G suit and positive pressure breathing). Estimation of head-level blood pressure and EEG are the two most common physiological metrics of acceleration stress (Lewis, McGovern, Miller, Eddy, & Forster, 1987). AGARD publication AGARD-LS-202 (1995) provides more detail on the physiological impact of Gz and the techniques of anti-G protective systems.

3.1.5.3 Effect on Performance

Other than a loss of vision with increasing Gz, there does not appear to be a linear relationship between the level of Gz stress and significant decrements in cognitive performance. However, very little is known about the effect of sustained high Gz exposure on cognitive capabilities. Individuals often experience what appears to be an instantaneous transition from complete consciousness to unconsciousness as Gz stress (either level or duration) reaches a certain level. Even in the complete absence of vision at moderate Gz levels, auditory and speech capabilities are maintained. Chelette, Albery, McCloskey, and Goodyear (1998) have shown that well-trained and well-protected participants can maintain performance on primary and secondary cognitive tasks throughout repeated high Gz exposures. The highly non-linear relationship between the magnitude of Gz and the resistance to blood flow in the cerebral drainage may account for these sharp transitions (Cirovic et al., 2000; Cirovic et al., 2001; Cirovic et al., 2002). However, a number of studies have detected decrements in arithmetic tasks (Frankenhauser, 1949), reaction times (Canfield, Comrey, & Wilson, 1948), memory tasks (Chambers & Hitchcock, 1963), tracking performance

(Albery, 1989), time estimation tasks (Frazier, Repperger, & Popper, 1990), and weight estimation tasks (Darwood, Repperger, & Goodyear, 1990).

3.1.5.4 Assessment Methods

Estimations of head-level blood pressure and EEG are the two most common physiological metrics of acceleration stress along with EMG recordings to quantify the straining effort. Ear opacity and pulse wave delay are often used to estimate head-level blood pressure, although more direct measurement of blood pressure has been performed in tactical aircraft using oscillometric tonometry of the temporal artery. Although these methods do not provide a direct measure of cerebral blood pressure, the evidence that the decreased transmural pressure in the extracranial veins of the neck is the critical factor in GLOC (Cirovic et al., 2000; Cirovic et al., 2001; Cirovic et al., 2002) would indicate that these metrics have strong face validity. The measurement of cerebral blood flow with Doppler ultrasound probes and the measurement of cerebral blood and tissue oxygen levels with near-infrared spectroscopy (NIRS; see NIRS section of report) are the best techniques for monitoring the physiological impact of Gz stress. Cerebral blood flow is difficult to monitor in the Gz environment since probe positioning is critical (Balldin, 1995) and NIRS technology suitable for Gz research has only recently become available. The Gz level, the length of time at high Gz, and the number and duration of repeated Gz exposures are also highly correlated with the likelihood of loss-of-consciousness.

Cognitive metrics are difficult to obtain during Gz exposures since the duration of each exposure is on the order of seconds. Cognitive tasks requiring substantial time to obtain sufficient data for statistical reliability and validity cannot be used. There are efforts underway to develop an acceleration performance assessment simulation system (A-PASS; O'Donnell, Cardenas, & Eddy, 1996) specifically for centrifuge research, incorporating various cognitive tasks that are required of the tactical pilot and that provide a measure of the operational impact of the high Gz environment. The continuous tracking tasks described above are the most popular in providing useful information as to any cognitive deficits during Gz, since they provide continuous time-series data that can be correlated with the continuous Gz time-series data. There have been several studies on the time required for cognitive function to recover to normal following GLOC incidents (Burns, 1995b). Complete amnesia of the final few seconds leading up to the events resulting in GLOC is common in both pilots and centrifuge subjects. Although several studies have shown no decrement in the performance of well-learned tasks once centrifuge participants have recovered full consciousness, there are numerous reports of emotional disturbances, feelings of not being one's self, and an excessive sense of embarrassment often lasting for hours after a GLOC occurrence.

3.1.5.5 References

AGARD (1995). *Current Concepts on G-Protection Research and Development* (AGARD-LS-202). Neuilly-sur-Seine, France: NATO Advisory Group for Aerospace Research and Development.

Albery, W.B. (1989). The Effects of Sustained Acceleration upon Human Operations. *Aviation, Space, and Environmental Medicine*, *60*, 943-948.

Ballidin, U.I. (1995). New Advances in Physiological Measurements during High-G: Technology. In *Current Concepts on G-Protection Research and Development* (AGARD-LS-202). Neuilly-sur-Seine, France: NATO AGARD.

Burns, J.W. (1995a). G-Protection Basis/Acceleration Physiology. In *Current Concepts on G-Protection Research and Development* (AGARD-LS-202). Neuilly-sur-Seine, France: NATO AGARD.

Burns, J.W. (1995b). G-Protection Capabilities and Current G-Protection Issues. In *Current Concepts on G-Protection Research and Development* (AGARD-LS-202). Neuilly-sur-Seine, France: NATO AGARD.

Burton, A.C. (1965). *Physiology and Biophysics of the Circulation*. Chicago, IL: Year Book Medical Publishers.

Canfield, A.A., Comrey, A.L., & Wilson, R.C. (1948). *The Effect of Increased Acceleration upon Human Abilities* (Report No. R-R.4). Los Angeles, CA: University of Southern California.

Chambers, R.M., & Hitchcock, L. (1963). *Effects of Acceleration on Pilot Performance* (NADC-MA-6110, DTIC-AD-408686). Warminster, PA: Naval Air Development Center.

Chelette, T.L., Albery, W.B., McCloskey, K.A., & Goodyear, C.D. (1998). Task Performance throughout Prolonged High G Exposure. In *Neurological Limitations of Aircraft Operations: Human Performance Implications* (pp. 3.1-3.4, AGARD-CP-579). Neuilly-sur-Seine, France: NATO AGARD.

Cirovic, S., Walsh, C., & Fraser, W.D. (2000). A Mathematical Model of Cerebral Perfusion Subjected to Gz Acceleration. *Aviation, Space, and Environmental Medicine*, *71*, 514-521.

Cirovic, S., Walsh, C., & Fraser, W.D. (2001). A Mechanical Model of Cerebral Circulation during Sustained Acceleration. *Aviation, Space, and Environmental Medicine*, *72*, 704-712.

Cirovic, S., Walsh, C., Fraser, W.D., & Gualino, S. (2003). The Effect of Posture and Positive Pressure Breathing on the Hemodynamics of the Internal Jugular Vein. *Aviation, Space, and Environmental Medicine*, *74*, 125-131.

Darwood, J.J., Repperger, D.W., & Goodyear, C.D. (1990). Mass Discrimination under +Gz Acceleration. *Aviation, Space, and Environmental Medicine*, *62*, 319-324.

Frankenhauser, M. (1949). Effects of Prolonged Gravitational Stress on Performance. *Acta Physiologica Scandinavia*, *104*, 10-11.

Frazier, J.W., Repperger, D.W., & Popper, S.E. (1990). Time Estimating Ability during +Gz Stress. *Aviation, Space, and Environmental Medicine*, *61*, 449.

Lewis, N., McGovern, J.B., Miller, J.C., Eddy, D.R., & Forster, E.M. (1987). EEG Indices of G-Induced Loss of Consciousness (G-LOC), pp. 29-1 to 29-12. In *AGARD Conference Proceedings No. 432 on Electrical and Magnetic Activity of the Central Nervous System: Research and Clinical Applications in Aerospace Medicine*. Neuilly-sur-Seine, France: NATO AGARD.

O’Donnell, R.D., Cardenas, R., & Eddy, D. (1996). *Assessing the Performance Impact of G-Forces: Design of the Accelerator-Performance Assessment Simulation System (A-PASS)* (AL/CF-TR-1996-0093). Wright-Patterson Air Force Base, OH: Air Force Material Command, WPAFB.

Prior, A.R.J. (1995). Positive Pressure Breathing for G Protection (PBG). In *Current Concepts on G-Protection Research and Development* (AGARD-LS-202). Neuilly-sur-Seine, France: NATO AGARD.

Rushmer, R.F., Beckman, E.L., & Lee, D. (1947). Protection of the Cerebral Circulation by the Cerebrospinal Fluid under the Influence of Radial Acceleration. *American Journal of Physiology*, *151*, 355-365.

Wood, E., Lambert, E., Baldes, E., & Code, C. (1946). Effects of Acceleration in Relation to Aviation. *Federal Proceedings*, *5*, 327-344.

3.1.6 Thermal Stress

3.1.6.1 Definition and Measurement

Thermal stress may be manifested as heat stress or cold stress. Heat stress results in an increase in blood flow to the body periphery to enhance cooling along with an increase in sweat gland activity for evaporative cooling. This response can sometimes lead to dehydration and reduced blood flow to the brain. Whereas it may be argued that environments in which heat is still a problem are not as frequently encountered as in the past, the commission of human errors due to the heat may still prove catastrophic in terms of human and monetary cost. Therefore, it becomes apparent that consideration of complex mental performance in hot environments is very important for the safety of operators and the systems within which they operate.

3.1.6.2 Background

Numerous studies have investigated the effects of heat stress on simple mental performance. Although many of these studies have reported some form of performance decrement (Iampietro et al., 1972; Mortagy & Ramsey, 1973; Pepler, 1958; Ramsey, Dayal, & Ghahramani, 1975; Wing & Touchstone, 1965), other studies have indicated that performance remains unaffected (Bell, Provins, & Hiorns, 1964; Chiles, 1958; Colquhoun, 1969; Nunneley, Dowd, Myhre, Stribley, & McNee, 1979), or even improves upon exposure to hot environments (Colquhoun & Goldman, 1972; Lovingood, Blyth, Peacock, & Lindsay, 1967; Poulton & Kerslake, 1965; Nunneley et al., 1979). In contrast to the availability of information on simple performance, fewer studies have examined the effects of heat on time-sharing performance. With respect to dual-task performance, the classic studies used a dual-task paradigm consisting of a task presented in the central visual field (tracking or choice reaction time) along with a peripheral light-detection task (Azer, McNall, & Leung, 1972; Bursill, 1958; Poulton, Edwards, & Colquhoun, 1974; Provins & Bell, 1970). With this paradigm, a progressive “funneling” of attention with increasing temperature was observed. This funneling is characterized by an increasing proportion of signals missed in the peripheral visual field compared to the proportion missed in the central visual field.

3.1.6.3 Effect on Performance

Ramsey (1983, 1995) provided reviews of the effects of heat and cold on human performance. An analysis of fifteen studies of the effects of heat stress on mental performance led to the setting by NIOSH (U.S. National Institute of Occupational Safety and Health) of an upper limit of exposure in terms of heat level and exposure duration (NIOSH, 1972). A later review of 22 studies concluded that a single temperature-time curve could not accurately represent the upper limit for unimpaired mental performance because of the large number of intervening variables, including the type of task (Ramsey & Morrissey, 1978). As a function of task category, the temperature-time performance effects exhibited one of two basic patterns. For reaction time and other simple mental tasks involving memory or speeded decision making, increases in either temperature or exposure duration increased the likelihood of impaired performance. For tracking, vigilance, and complex tasks, all of which require sustained attention, increases in temperature degrade performance more than increases in exposure time (Ramsey & Morrissey, 1978). Due to the complexity of the problem, the most recent NIOSH criteria do not address the issue of mental performance limitations under heat stress (NIOSH, 1986).

Hancock and Vasmatzidis (in press) provide a review of the current state of knowledge regarding the effects of heat stress on cognitive performance. Based on their review and on the development of a theoretical model for heat stress effects, they propose a new attentional resource approach defining temperature-exposure duration thresholds as parallel lines when plotting temperature vs. the logarithm of time.

With respect to complex time-sharing performance (i.e., performance involving three or more concurrent tasks), studies are relatively scarce. In one such study, Iampietro, Chiles, Higgins, and Higgins (1969) found that horizontal tracking performance (when combined with a monitoring and mental arithmetic task) and mental arithmetic performance (when combined with a monitoring task) were significantly lower during a 30-minute exposure at 71°C (160°F) compared to a 15-minute pre-exposure period. In a similar study, Chiles, Iampietro, and Higgins (1972) combined tracking with three monitoring tasks and with mental arithmetic and monitoring tasks. They found that tracking efficiency declined significantly during a 15-minute exposure at 35°C compared to performance in a thermally neutral environment. Vasmatzidis, Schlegel, and Hancock (2002) evaluated performance on three dual-task combinations (display monitoring with mathematical processing, memory search with mathematical processing, and unstable tracking with memory search) and on a multiple-task (SYNTASK) for two hours in each of six climates. The climates were obtained by generating each of three wet bulb globe temperature (WBGT) temperatures (22°C, 28°C and 34°C) with two relative humidity levels (30% and 70%). There was a significant heat stress effect on display monitoring and unstable tracking performance and on the SYNTASK visual monitoring and auditory discrimination tasks. Additionally, at 34°C WBGT, 70% relative humidity was more detrimental to performance than 30% relative humidity.

3.1.6.4 Assessment Methods

Monitoring the thermal properties of the operator's environment is an important element of protecting the operator from the effects of heat stress or cold stress. Exposing operators to heat stress may lead to heat strain with physiological symptoms ranging from muscle cramps to heat stroke. Determining which parameters of the thermal environment should be measured, how to measure them, and how to use the resulting information is an important component of OFS assessment in extreme thermal environments.

Several parameters of the thermal environment are readily measurable and may be combined in various ways to provide a heat stress index, a single indicator of the severity of the heat stress environment. Information on specific instruments to measure these environmental parameters and on placement of the instruments at the work site is available from a number of sources, including standards documents (ASHRAE Standard 55, 1992; ISO Standard #7726, 1985), instrument manufacturers, and various journal articles. The most important parameters include air temperature (dry-bulb temperature), air movement (wind speed and air velocity), relative humidity (wet-bulb and dry-bulb temperatures linked using a psychrometric chart), and mean radiant temperature (non-ionizing radiation primarily in the infrared region). Reducing the work site temperature and humidity, increasing the air movement, and shielding or shading the work area to reduce the amount of radiation can reduce the risk of heat strain and the accompanying symptoms.

3.1.6.4.1 Dry-Bulb Temperature

Dry-bulb temperature is what is commonly referred to as air temperature. Dry-bulb temperature may be measured using a common mercury-in-glass thermometer or an electronic instrument with a thermistor or thermocouple sensor. The sensor should be placed in the open air in a location that is shaded from the sun and shielded from other radiation sources like furnaces. Although dry-bulb temperature helps indicate the direction and amount of convective heat exchange with the human body, by itself it is a poor indicator of heat stress.

3.1.6.4.2 Wet-Bulb Temperature

Wet-bulb temperature varies with the relative humidity and is measured by attaching a wetted cotton wick (or sock) to a thermometer or other temperature sensor. Evaporation of water from the sock cools the thermometer. The amount of cooling depends on the humidity and movement of the surrounding air. Relative humidity describes the water vapor pressure at a given temperature as a percentage of the saturated water vapor pressure at that temperature. When relative humidity is 100% and the air is completely saturated, the wet-bulb and dry-bulb temperatures are equal. When they are not equal, the relative humidity can be determined from a psychrometric chart using the dry-bulb and wet-bulb readings. There are two types of wet-bulb temperatures, aspirated wet-bulb and natural wet-bulb.

Aspirated Wet-Bulb Temperature

The aspirated or psychrometric wet-bulb temperature is obtained by using a fan or by twirling dry-bulb and wet-bulb thermometers mounted on a sling arm (sling psychrometer) so that air is forced over the wetted wick at a speed greater than 3 m/s. Psychrometric wet-bulb temperature is used with a psychrometric chart to determine relative humidity based on the difference between the two thermometer readings. Besides the sling psychrometer, electronic instruments are available which provide direct readings of wet-bulb and dry-bulb temperatures, relative humidity, and dew point.

Natural Wet-Bulb Temperature

The natural wet-bulb temperature is obtained by exposing the wet-bulb thermometer with the wetted cotton wick to the natural, or prevailing, air movement. Because the evaporative cooling of the wick depends on the environmental conditions at the work site, the natural wet-bulb temperature is a good indicator of the ability of the surrounding environment to support body cooling.

3.1.6.4.3 Psychrometric Chart

The psychrometric chart shows the relationship between dry-bulb temperature, wet-bulb temperature, and humidity. Two axes plot the adjusted dry-bulb temperature (mean of air temperature and radiant temperature) and the aspirated wet-bulb temperature. Other axes provide the relative humidity and the absolute humidity. The chart also indicates water vapor pressure, which is used to determine the maximum amount of evaporative cooling that can be supported by the environment.

3.1.6.4.4 Globe Temperature

Globe temperature is used to estimate the mean radiant temperature, the average temperature of the solid surroundings. Globe temperature is measured using a temperature sensor (thermometer or thermistor) in the middle of a matte black-painted copper sphere. Originally the diameter of the sphere was 6 inches, but new instruments use a smaller sphere. Radiant heat from the sun or other hot objects is absorbed by the sphere and heats the thermometer.

3.1.6.4.5 Air Movement

Outdoor air movement or wind speed is measured by a mechanical cup or propeller anemometer. Indoor air movement at low velocities is measured by a hot-wire or heated-bead anemometer. Air speed affects the amount of heat transferred to or from the body due to a difference between body skin temperature and air temperature. It also affects the rate of evaporative cooling.

3.1.6.4.6 Wet-Bulb Globe Temperature

The WBGT index is a weighted average of the natural wet-bulb temperature, the globe temperature, and the dry-bulb temperature (if outside in direct sunlight). Although each of these temperatures may be measured separately, commercial heat stress monitors exist which calculate the WBGT index from the three temperatures and display each of the measurements.

3.1.6.5 References

American Society of Heating, Refrigerating and Air-Conditioning Engineers (1992). *Thermal Environmental Conditions for Human Occupancy* (ASHRAE Standard 55). Atlanta, GA: ASHRAE.

Azer, N.Z., McNall, P.E., & Leung, H.C. (1972). Effects of Heat Stress on Performance. *Ergonomics*, *15*, 681-691.

Bell, C.R., Provins, K.A., & Hiorns, R.W. (1964). Visual and Auditory Vigilance during Exposure to Hot and Humid Conditions. *Ergonomics*, *7*, 279-288.

Bursill, A.E. (1958). The Restriction of Peripheral Vision during Exposure to Hot and Humid Conditions. *Quarterly Journal of Experimental Psychology*, *10*, 113-129.

Chiles, W.D. (1958). Effects of Elevated Temperature on Performance of a Complex Mental Task. *Ergonomics*, *2*, 89-96.

Chiles, W.D., Iampietro, P.F., & Higgins, E.A. (1972). Combined Effects of Altitude and High Temperature on Complex Performance. *Human Factors*, *14*, 161-172.

Colquhoun, W.P. (1969). Effects of Raised Ambient Temperature and Event Rate on Vigilance Performance. *Aerospace Medicine*, *40*, 413-417.

Colquhoun, W.P., & Goldman, R.F. (1972). Vigilance under Induced Hyperthermia. *Ergonomics*, *15*, 621-632.

Hancock, P.A., & Vasmatzidis, I. (in press). Effects of Heat Stress on Cognitive Performance: The Current State of Knowledge. *International Journal of Hyperthermia*.

Iampietro, P.F., Chiles, W.D., Higgins, E.A., & Higgins, H.L. (1969). Complex Performance during Exposure to High Temperatures. *Aerospace Medicine*, *40*, 1331-1335.

Iampietro, P.F., Melton, C.E., Higgins, E.A., Vaughan, J.A., Hoffman, S.M., Funkhouser, G.E., & Saldivar, J.T. (1972). High Temperature and Performance in a Flight Task Simulator. *Aerospace Medicine*, *43*, 1215-1218.

International Standards Organization. (1985). *Thermal Environments – Instruments and Methods for Measuring Physical Quantities* (ISO Standard #7726). Geneva: ISO.

Lovingood, B.W., Blyth, C.S., Peacock, W.H., & Lindsay, R.B. (1967). Effects of *d*-amphetamine Sulfate, Caffeine and High Temperature on Human Performance. *Research Quarterly*, *38*, 64-71.

Mortagy, A.K. & Ramsey, J.D. (1973). Monitoring Performance as a Function of Work/Rest Schedule and Environmental Stress. *American Industrial Hygiene Association Journal*, *34*, 474-480.

National Institute for Occupational Safety and Health. (1972). *Criteria for a Recommended Standard – Occupational Exposure to Hot Environments* (USGPO-HSM 72-10269). Washington, DC: U.S. Government Printing Office.

National Institute for Occupational Safety and Health. (1986). *Occupational Exposure to Hot Environments: Revised Criteria* (USGPO-HSM 86-113). Washington, DC: U.S. Government Printing Office.

Nunneley, S.A., Dowd, P.J., Myhre, L.G., Stribley, R.F., & McNee, R.C. (1979). Tracking-Task Performance during Heat Stress Simulating Cockpit Conditions in High-Performance Aircraft. *Ergonomics*, *22*, 549-555.

Pepler, R.D. (1958). Warmth and Performance: An Investigation in the Tropics. *Ergonomics*, *2*, 63-88.

Poulton, E.C., & Kerslake, M.B. (1965). Initial Stimulating Effect of Warmth upon Perceptual Efficiency. *Aerospace Medicine*, *36*, 29-32.

Poulton, E.C., Edwards, R.S., & Colquhoun, W.P. (1974). The Interaction of a Loss of a Night's Sleep with Mild Heat: Task Variables. *Ergonomics*, *17*, 59-73.

Provins, K.A., & Bell, C.R. (1970). Effects of Heat Stress on the Performance of Two Tasks Running Concurrently. *Journal of Experimental Psychology*, *85*, 40-44.

Ramsey, J.D. (1983). Heat and Cold. In G.R.J. Hockey (Ed.), *Stress and Fatigue in Human Performance* (pp. 33-60). Chichester, England: Wiley.

Ramsey, J.D. (1995). Task Performance in Heat: A Review. *Ergonomics*, *38*, 154-165.

Ramsey, J.D., Dayal, D., & Ghahramani, B. (1975). Heat Stress Limits for the Sedentary Worker. *American Industrial Hygiene Association Journal*, *36*, 259-265.

Ramsey, J.D., & Morrissey, S.J. (1978). Isodecrement Curves for Task Performance in Hot Environments. *Applied Ergonomics*, *9*, 66-72.

Vasmatzidis, I., Schlegel, R.E., & Hancock, P.A. (2002). An Investigation of Heat Stress Effects on Time-Sharing Performance. *Ergonomics*, *45*, 218-239.

Wing, J.F., & Touchstone, R.M. (1965). *The Effect of High Ambient Temperature on Short-Term Memory* (Report AMRL-TR-65-103). Wright-Patterson AFB, OH: Aerospace Medical Research Laboratories.

3.2 INDIVIDUAL STATE

3.2.1 Circadian Rhythms

3.2.1.1 Definitions and Measurement

The circadian system status of a human organism is a universal condition of its general functional state (Alyakrinsky, 1989). Circadian Rhythms (Halberg, Halberg, Barnum, & Bittner, 1959) refer to cyclical body processes having a period of approximately one day (diurnal rhythm). Corresponding diurnal cycling is also evident in performance measures. Mathematically, rhythmic processes may be conveniently defined by four parameters: (1) the *period*, or time for completion of one cycle, (2) the *rhythm-adjusted mean*, (3) the *phase*, or relative location of the rhythm in time (with reference to the maximum or minimum value), and (4) the *amplitude*, or range of oscillation of the process (difference between maximum and minimum values).

The rhythm is circadian when the period is approximately 24 hours (between 20 and 28 hours), ultradian when the period is shorter than 20 hours, and infradian when the period is longer than 28 hours. If a rhythm can be approximated by a cosine curve model (Halberg, Carandente, Cornelissen, & Katinas, 1977), the midline estimating statistic of rhythm (MESOR) represents the value midway between the highest and the lowest values of the function used to approximate the rhythm. The amplitude of a rhythm is defined as one-half the difference between the highest and the lowest point of the mathematical model. Using this mathematical model, the location in time of the rhythm is defined by the highest point (peak or acrophase) or the lowest point (trough or bathyphase) of the variable in relation to a phase reference chosen by the investigator (e.g., local midnight for a circadian rhythm).

Two important features of circadian rhythms are that they seem to be genetically determined (Aschoff & Wever, 1981) and that they are continuously modulated, modified, and adjusted in time (entrained or synchronized) by periodic events in the environment (*zeitgebers* or synchronizers; Pittendrigh, 1981). The genetic origin of the rhythms suggests the existence of a "biological clock" (Miller, Morin, Schwartz, & Moore, 1996) located in a cerebral structure (paired suprachiasmatic nuclei). The clock acts as a pacemaker for numerous endogenous rhythms in the absence of external time cues, but is sensitive to information provided by synchronizers allowing it to be continuously reset. In humans, *zeitgebers*/synchronizers include the daily light/dark cycle, social routine, noise/silence periods, activity/rest schedules, and timing of food intake.

3.2.1.2 Background

Research in circadian rhythms has focused on chronobiology (i.e., the cycling of physiological processes over time) and cyclical variations in performance. Of particular interest are variations in performance based on time of day, phase shifts associated with shift work or time zone displacements, or isolation from the normal zeitgeber.

3.2.1.3 Effects on Performance

Hockey provides a summary of research on diurnal variations in performance and adaptation to altered schedules (Hockey, 1986). A substantial amount of research over the past four decades confirms the existence of circadian variation in a range of performance tasks. Research has also demonstrated that different kinds of tasks have different rhythms. For example, acoustical reaction time is fastest at 6 p.m. and slowest at 3 a.m. (Voigt, Engel, & Klein, 1968), while error frequency peaks at 3 a.m. and at 3 p.m. (Bjerner & Swensson, 1953). It has been shown that physiologic parameters (pupillometric variables and sleep latencies) show peak levels of alertness at 7:00 a.m., while self-assessment scales and simple cognitive test performance have demonstrated peaks around noon (ranging from 11:00 a.m. to 3:00 p.m.; Kraemer et al., 2000). For most tasks, performance is generally better later in the working day (normal day shift) than at the beginning of the day, although tasks involving immediate memory often show the opposite effect (Colquhoun, 1971). Thus, the nature of the processing requirements exerts a major influence on the form of the time-of-day effect. Folkard has shown that tasks involving both speeded processing and a high dependence on the use of immediate memory exhibit a time-of-day effect with a performance peak in the middle of the day, a compromise between the extremes shown for the respective individual tasks (Folkard, 1975). This compromise applies to tasks involving reasoning, complex arithmetic, or other combinations of speed stress with a heavy memory load. Performance decrements in a brief cognitive visual search task were evident on speed but not on accuracy indices, and were strictly related to the deterioration of oculomotor performance, indicating a clear circadian effect (De Genarro, Ferrara, Curcio, & Bertini, 2001).

Demonstrated results need to account for the human biorhythmic profile (Halberg & Halberg, 1980) because of different acrophases for "morning", "evening" and arrhythmic types of profiles (i.e., the fastest visual reaction time for these types is at 12 noon, 10 p.m., and 12 noon or 6 p.m., respectively; Doskin & Kuindzhy, 1989). Accordingly, performance effectiveness depends on the correspondence of time and the method of measuring the human biorhythmic type.

According to the Gubin's concept of the circadian organization of living systems, all amplitude-phase relations are changed in ontogenesis (Gubin & Gerlovin, 1980). Hence, when we measure human psychophysiological parameters in relation to circadian rhythms we must correct for age.

Adaptation to shifts in phase resulting from shift work or time zone displacement is easily identified by observing shifts in body temperature rhythm. The extent of the adaptation is a function of the degree of displacement, the length of time allowed, social patterns, and individual differences. Few studies have measured performance changes during circadian rhythm adjustment. Folkard and his colleagues demonstrated that cognitive tasks having high requirements for internal processing, working memory, decision making, verbal reasoning, and arithmetic adapt more rapidly than visual search and manual dexterity tasks (Folkard, Knauth, & Monk, 1978; Folkard, Monk, & Lobban, 1978).

Shift work and jet lag can disrupt circadian rhythms, producing detrimental effects on alertness, performance, and sleep. The human circadian timekeeping system responds to changes in work schedule (Monk, 1990; Monk, Folkard, & Wedderburn, 1996). It has been established that shiftwork tolerance is connected with introversion, neuroticism, morningness, control of behavioral arousal, and parameters of circadian rhythms. The most important predictors of shiftwork tolerance are the dimensions of control of behavioral arousal and morningness, while the most important criterion is sleep quality (Petz & Vidacek, 1999). Current opinion is that temperature is the most stable indicator of circadian influence on human performance. The relationship between temperature and performance is stronger in younger rather than in older subjects, and particularly weak in older men (Monk & Kupfer, 2000).

Synchronization of the circadian rhythms in the organism can be a criterion of well-being and capacity. However, circadian rhythm disruption is an index of desynchronization and an infringement on the adaptation, which can be further provoked by disturbance of the sleep-wake balance (Stepanova, 1986). It has been shown that if electric lighting, as currently employed, contributes to “circadian disruption”, it may be an important cause of “endocrine disruption” and thereby contribute to a high risk of breast cancer in industrialized societies (Stevens & Rea, 2001). Antoniadis, Ko, Ralph, & McDonald (2000) have pointed out that in human beings and animal models, cognitive performance is often impaired in natural and experimental situations where circadian rhythms are disrupted. This observation includes a general decline in cognitive ability and fragmentation of behavioral rhythms in the aging population of numerous species.

The adaptation of circadian rhythms after drastic lagging of the sleep-wake rhythm occurs due to the rhythm period lengthening (Mills, 1977; Minors, Mills, & Waterhouse, 1977). In general, adaptation to the lagging of the sleep-wakefulness rhythm, to the shorter and longer day, does not occur. In contrast to day-shift work, it has been shown that performance errors by night-shift workers typically occur in the early morning (Nakano et al., 2000).

Fluctuations of circadian rhythms induced by isolation from external *zeitgebers* have been revealed in a number of studies and have demonstrated circadian rhythm lengthening (Aschoff, Hoffman, Pohl, & Wever, 1975; Aschoff & Wever, 1976). Subjects in isolation can show disturbances of sleep, mood, and vigilance if their biological rhythms run “out of phase” (Zulley, 2000). Akoev has shown that increased stress from conditions of being in “isolation from time” can be eliminated in the organism by slowing down the biological rhythmicity (Akoev, 1976).

3.2.1.4 Assessment Methods

Several parameters that are subject to circadian rhythmicity are not easily entrained by *zeitgebers*. They are easily measured and are thus considered strong markers. These “gold standards” include body temperature, cortisol, catecholamines and/or melatonin plasma or salivary levels, urinary volume, and sweat electrolytes. Other variables such as sleep/wakefulness state, heart rate, heart rate variability, and blood pressure, although more easily synchronized, have been measured because they are linked to fatigue level and are easy to measure. Nevertheless, an accurate and reliable description of circadian rhythms demands repeated measurement of each variable at least once per hour.

3.2.1.4.1 Core Temperature

Body core temperature is a strong marker of circadian rhythms (peak at 5:00 p.m., trough at 5:00 a.m.) that can be continuously measured during field studies by telemetry (Cor-Temp System, Human Technologies Inc., Florida, USA). Each subject is asked to ingest one transmitting capsule, and data are collected by means of a receiving antenna. Absolute temperature values can be averaged over 1-hr to 2-hr periods for each subject.

This method of telemetry over a prolonged period could potentially induce variations in absolute values (artifacts) at the time of change of the capsule. Consequently, relative values of temperature, as well as of hormonal samples, can be calculated as the difference between the absolute data and a baseline value representative of a given time period (Beaumont et al., 2001; Lagarde et al., 2000).

3.2.1.4.2 Hormonal Sampling

Within subjects, predominantly negative correlations have emerged between good performance and high plasma levels of cortisol and melatonin. Plasma cortisol and melatonin can be easily measured from salivary samples during field studies. An example of such a study of jet lag was conducted by American and French scientists to investigate the potential caffeine-induced resynchronization of endogenous melatonin and cortisol secretions (Piérard et al., 2001).

3.2.1.4.3 Subjective Measures of Sleep and Alertness

Sleep/wakefulness and alertness, and subsequently cognitive performance, follow a circadian rhythm (Lavie, 2001): sleep propensity is maximal when core temperature begins to decrease and vice-versa for wakening. Rapid Eye Movement (REM) sleep occurs at the end of the night and alertness exhibits a biphasic pattern during the nycthmeron, with two hypovigilance periods (2:00-6:00 a.m.; 2:00-6:00 p.m.) and two hypervigilance periods (9:00-12:00 a.m.; 7:00-10:00 p.m.).

Sleep and somnolence can be subjectively determined with high reliability using sleep questionnaires/ diaries and sleepiness scales respectively. Whereas sleep questionnaires and logs are numerous, the Stanford Sleepiness Scale (SSS) has been the standard measure of introspective sleepiness for many years (Kraemer et al., 2000; Patat et al., 2000). Participants are asked to choose one of seven statements to describe the self-assessed current state. Advantages of the SSS include its brevity, ease of administration, and the fact that it can be completed repeatedly, which is useful for evaluating circadian rhythm influences on sleepiness. At the opposite extreme, the Epworth Sleepiness Scale (ESS), which describes the drive to sleep rather than sleepiness (Johns, 1991), is of questionable value when re-administered within a brief time interval. In addition, the sensitivity of the ESS to age, acute sleep disturbance or loss, and drugs is not known.

3.2.1.4.4 Secondary Measures

Electroencephalography (EEG)

Sleep and sleepiness are objectively measured using EEG during clinical and research studies in the laboratory and during field studies with continuous EEG recordings. Sleep is analyzed from standard polysomnographic recordings that include electroencephalography (EEG), electrooculography (EOG) of each eye (oblique and horizontal derivations), and chin electromyography (EMG). EEG signals are recorded from electrodes attached to the scalp with collodion or held in position with a special electrode cap (Beaumont et al., 2000; Beaumont et al., 1996). Five sites, two over central scalp areas (C3/Cz) and two over occipital scalp areas (O1/O2), referenced to the left ear (A1) are sufficient to get a good evaluation of sleep in healthy subjects, even during field studies. After amplification and filtering, all signals can be either directly read on a computer or stored using a portable recorder to be analyzed later according to the standard criteria (Rechtschaffen & Kales, 1968).

Sleepiness can also be determined from continuous EEG recordings by directly scoring microsleep episodes (Lagarde et al., 2000). These events are shown by increased amounts of alpha and theta band activity in behaviorally awake humans who have been deprived of sleep. The evidence suggests that these microsleep episodes are indicants of sleepiness. To precisely determine the degree of sleepiness, it is useful to accumulate all microsleep episodes over a given period.

Sleepiness is classically quantified from intermittent EEG recording using the Multiple Sleep Latency Test (MSLT). This method is based on the assumption that sleepiness is a physiological need state that leads to an increased tendency to fall asleep (Carskadon & Dement, 1977). This test measures, at 2-h intervals throughout the day, the latency to fall asleep while lying with eyes shut, in a quiet, dark room. Individuals undergoing an MSLT are instructed to allow themselves to fall asleep or not to resist falling asleep. Sleep latencies in healthy normal individuals range from 10 to 20 minutes. Sleepiness is defined as a mean sleep latency of less than 5 to 6 minutes. However, this method is of limited use due to the fact that subjects are not permitted to remain in bed between nap test sessions. This can disturb the recovery sleep that is scheduled at a given time. Moreover, subjects should not engage in vigorous pre-test activity because it will alter the test outcome. The room must be dark and quiet during testing, and polysomnographic recordings must be made during the nap opportunities.

Wrist Actigraphy

Wrist actigraphy is used as an objective but indirect criterion of alertness (Brown, Smolenski, D'Alonzo, & Redman, 1990). This method is based on the fact that during sleep there is little movement whereas during wakefulness there is increased movement. Subjects are asked to wear a piezoelectric accelerometer on their non-dominant wrist throughout the operation. The number of movements with acceleration greater than 0.1 G are classically sampled, stored in 1-min epochs, and often averaged hour by hour. The collected data are examined for activity versus inactivity and analyzed for wake versus sleep.

Actigraphy is a very useful tool, especially during field studies, to determine circadian rhythms and sleep/wake cycles. Many studies have been conducted in shift workers, in-flight crew, and in persons with jet lag (Lowden & Akerstedt, 1999; Monk, Buysse, & Rose, 1999).

Cardiovascular Variables

Time of day has a significant effect on cardiac reactivity (CR), and two types of activity – sleep and arousal – have a non-similar influence (Lemmer, 1989). Measurements of heart rate (HR) and heart rate variability (HRV) are increasingly used as markers of CR (Lemmer, 1989). To examine circadian variation in HR and HRV, mean RR interval and frequency domain indices (very low, low, and high frequency indices) are determined hourly. A chronobiological analysis can be made using the cosinor method or spectral (periodogram) analysis (Smolensky et al., 1976).

A significant circadian variation in HR and HRV is present from late infancy or early childhood, and is characterized by a rise in variability during sleep (Cornélissen et al., 1990; Cornélissen et al., 1988). On the other hand, the low-frequency to high-frequency ratio increases during the daytime. The appearance of these circadian rhythms has been associated with sleep maturation (Maquet et al., 1996). The time of peak variability does not depend on age. These data confirm a progressive maturation of the autonomic nervous system and support the hypothesis that the organization of sleep, associated with sympathetic withdrawal, is responsible for these rhythms (Mancia, 1993).

The heart rate-blood pressure (HR-BP) product is the strongest correlate of short-term activity. The HR of working people peaks at 8:00-9:00 p.m. with a trough at 8:00-9:00 a.m., while BP is higher during the day, intermediate in the evening, low during the night, and rises before awakening (Halberg, Halberg, & Shankaraiah, 1981; Smolensky, Halberg, & Sargent, 1972). The 24-h profile of BP, as observed under normal circumstances, results from an endogenous circadian component rather than from environmental and behavioral factors such as the occurrence of sleep (Halberg, Good, & Levine, 1966). Thus, although changes in posture usually contribute to the extent of within-day change, circadian rhythmicity persists with statistical significance under conditions of bed-rest (Stadick, Bryans, Halberg, & Halberg, 1988).

Respiration Rate

Respiration rate has a peak at 4 p.m. and a trough at 4 a.m. during a normal day schedule, while conspicuous changes occur in the regulation of breathing across the behavioral stages of sleep (Barnes, 1985; Phillipson & Bowes, 1986). The transition from wakefulness to non-REM (NREM) sleep is characterized by a breathing instability during stages 1 and 2 of sleep, while regular breathing sets in with deep NREM sleep (stages 3 and 4; Trinder, Whitworth, Kay, & Wilkin, 1992). REM sleep is also characterized by a marked irregularity of the rhythm of breathing (Phillipson, 1978). No significant changes of respiratory patterns (and of circulatory parameters as well) have been revealed under conditions of sleep deprivation (Johnson, Slye, & Dement, 1965).

3.2.1.5 References

Akoev, I.G. (1976). Some Issues of Fundamental Research on Space Biophysics. In *Experimental Research on Space Biophysics* (pp. 7-30). Pushchino.

Alyakrinsky, B. (1989). The Problem of Circadianness. In O. Gazenko (Ed.), *Biorhythmic Researches in Space Biology and Medicine* (Vol. 64, pp. 12-34). Moscow: Nauka.

Antoniadis, E.A., Ko, C.H., Ralph, M.R., & McDonald, R.J. (2000). Circadian Rhythms, Aging and Memory. *Behavioral Brain Research*, *114*, 221-233.

Aschoff, J., Hoffman, K., Pohl, H., & Wever, R. (1975). Re-Entrainment of Circadian Rhythms after Phase-Shifts to the Zeitgeber. *Chronobiologia*, *2*, 23-78.

Aschoff, J., & Wever, R. (1976). Human Circadian Rhythms: A Multi-Oscillator System. *Federal Proceedings*, *35*, 2326-2332.

Aschoff, J., & Wever, R. (1981). The Circadian System of Man. In J. Aschoff (Ed.), *Handbook of Behavioural Neurobiology* (pp. 311-331). London: Plenum.

Barnes, P.J. (1985). Circadian Variation in Airway Function. *American Journal of Medicine*, *79* (Suppl 6A), 5-9.

Beaumont, M., Batéjat, D., Piérard, C., Coste, O., Doireau, P., Van Beers, P., Chauffard, F., Chassard, D., Enslen, M., Denis, J.B., & Lagarde, D. (2001). Slow Release Caffeine and Prolonged (64-h) Continuous Wakefulness: Effects on Vigilance and Cognitive Performance. *Journal of Sleep Research*, *10*, 265-276.

Beaumont, M., Denis, J.B., Piérard, C., Batéjat, D., Van Beers, P., Chauffard, F., French, J., & Lagarde, D. (2000). Slow Release Caffeine Alleviates Jet Lag-Related Sleep Perturbations and Daytime Sleepiness after a 7-time Zone Eastbound Flight. *Aviation, Space, and Environmental Medicine*, *71*, 298.

Beaumont, M., Goldenberg, F., Lejeune, D., Marotte, H., Harf, A., & Lofaso, F. (1996). Effect of Zolpidem on Sleep and Ventilatory Patterns at Simulated Altitude of 4,000 Meters. *American Journal of Respiratory and Critical Care Medicine*, *153*, 1864-1869.

Bjerner, B., & Swensson, A. (1953). Schlichtarbeit und rhytmus. *Acta Medica Scandinavica, Suppl 278*, 102-107.

Brown, A., Smolenski, M., D'Alonzo, G., & Redman, D. (1990). Actigraphy: A Means of Assessing Circadian Patterns in Human Activity. *Chronobiology International*, *7*, 125-133.

Carskadon, M., & Dement, W. (1977). Sleep Tendency: An Objective Measure of Sleep Loss. *Sleep Research*, *6*, 200.

Colquhoun, W.P. (1971). Circadian Variations in Mental Efficiency. In W.P. Colquhoun (Ed.), *Biological Rhythms in Human Performance* (pp. 39-107). London: Academic Press.

Cornélissen, G., Bakken, E., Delmore, P., Orth-Gomér, K., Akerstedt, T., Carandente, O., Carandente, F., & Halberg, F. (1990). From various Kinds of Heart Rate Variability to Chronocardiology. *American Journal of Cardiology*, *66*, 863-868.

Cornélissen, G., Scarpelli, P.T., Halberg, F., Halberg, J., Halberg, F., & Halberg, E. (1988). Cardiovascular Rhythms: Their Implications and Applications in Medical Research and Practice. In W.T.J.M. Hekkens, G.A. Kerkhof, & W.J. Rietveld (Eds.), *Trends in Chronobiology* (pp. 335-355). Oxford: Pergamon Press.

De Genarro, L., Ferrara, M., Curcio, G., & Bertini, M. (2001). Visual Search Performance across 40 h of Continuous Wakefulness: Measures of Speed and Accuracy and Relation with Oculomotor Performance. *Physiology and Behavior*, *74*, 197-204.

Doskin, V., & Kuindzhy, N. (1989). *Biological Rhythms of the Growing Organism*. Moscow: Nauka.

Folkard, S. (1975). Diurnal Variation in Logical Reasoning. *British Journal of Psychology*, *66*, 1-8.

Folkard, S., Knauth, P., & Monk, T.H. (1978). The Effect of Memory Load on the Circadian Variation in Performance Efficiency under a Rapidly Rotating Shift System. *Ergonomics*, *19*, 479-488.

Folkard, S., Monk, T.H., & Lobban, M.C. (1978). Short and Long-Term Adjustment of Circadian Rhythms in 'Permanent' Night Nurses. *Ergonomics*, *21*, 785-799.

Gubin, G.D., & Gerlovin, E.S. (1980). *Diurnal Rhythms of Biological Processes and Their Adaptive Meaning in Onto- and Phylogenesis*. Novosibirsk: Nauka.

Halberg, E., & Halberg, F. (1980). Chronobiologic Study Design in Everyday Life, Clinic and Laboratory. *Chronobiologia*, *7*, 95-120.

Halberg, E., Halberg, F., & Shankaraiah, K. (1981). Plexo-Serial Linear-Nonlinear Rhythmometry of Blood Pressure, Pulse and Motor Activity by a Couple in their Sixties. *Chronobiologia*, *8*, 351-366.

Halberg, F., Carandente, F., Cornelissen, G., & Katinas, G.S. (1977). Glossary of Chronobiology. *Chronobiologia, Suppl. 4*, 1-189.

Halberg, F., Good, R.A., & Levine, H. (1966). Some Aspects of the Cardiovascular and Renal Circadian System. *Circulation*, *34*, 715-717.

Halberg, F., Halberg, E., Barnum, C.P., & Bittner, J.J. (1959). Physiologic 24-h Periodicity in Human Beings and Mice, the Lighting Regimen and Daily Routine. In R.B. Withrow (Ed.), *Photoperiodism and Related Phenomena in Plants and Animals* (Vol. 55, pp. 803-878). Washington, DC: American Association for the Advances in Science.

Hockey, G.R.J. (1986). Changes in Operator Efficiency as a Function of Environmental Stress, Fatigue and Circadian Rhythms. In K. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. 2: Cognitive Processes and Performance, pp. 1-49). New York: Wiley.

Johns, M.W. (1991). A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep*, *14*, 540-545.

Johnson, L.C., Slye, E.S., & Dement, W.C. (1965). Electroencephalographic and Autonomic Activity During and After Prolonged Sleep Deprivation. *Psychosomatic Medicine*, *27*, 415-423.

Kraemer, S., DankerHopfe, H., Dorn, H., Schmidt, A., Ehlert, I., & Herrmann, W.M. (2000). Time-of-Day Variations of Indicators of Attention: Performance, Physiologic Parameters, and Self-Assessment of Sleepiness. *Biological Psychiatry*, *48*, 1069-1080.

Lagarde, D., Batéjat, D., Sicard, B., Trocherie, S., Chassard, D., Enslen, M., & Chauffard, F. (2000). Slow Release Caffeine: A New Response to the Effects of a Limited Sleep Deprivation. *Sleep*, *23*, 651-661.

Lavie, P. (2001). Sleep-Wake as a Biological Rhythm. *Annual Review of Psychology*, *52*, 277-303.

Lemmer, B. (1989). Circadian Rhythms in the Cardiovascular System. In J. Arendt, D.S. Minor, & J.M. Waterhouse (Eds.), *Biological Rhythms in Clinical Practice* (pp. 51-70). London: Butterworth.

Lowden, A., & Akerstedt, T. (1999). Eastward Long Distance Flights, Sleep and Wake Patterns in Air Crews in Connection with a Two-Day Layover. *Journal of Sleep Research*, *8*, 15-24.

Mancia, G. (1993). Autonomic Modulation of the Cardiovascular System during Sleep. *New England Journal of Medicine*, *328*, 347-349.

Maquet, P., Peters, J., Aerts, J., Delfiore, G., Degueldre, C., Luxen, A., & Franck, G. (1996). Functional Neuroanatomy of Human Rapid-Eye-Movement Sleep and Dreaming. *Nature*, *383*, 163-166.

Miller, J.D., Morin, L.P., Schwartz, W., & Moore, R.Y. (1996). New Insights into the Mammalian Circadian Clock. A State-of-the-Art Review. *Sleep*, *19*, 641-667.

Mills, J.N. (1977). Manipulation of Synchronizers. *Proceedings of International Union of Physiology and Sciences*, *12*, 779.

Minors, D.S., Mills, J.N., & Waterhouse, J.M. (1977). The Nature of the "Adaptation" of Human Circadian Rhythmicity to an Altered Time Schedule. *Chronobiologia*, *4*, 133.

Monk, T.H. (1990). The Relationship of Chronobiology to Sleep Schedules and Performance Demands. *Work Stress*, *4*, 227-236.

Monk, T.H., Buysse, D.J., & Rose, L.R. (1999). Wrist Actigraphic Measures of Sleep in Space. *Sleep*, *22*, 948-954.

Monk, T.H., Folkard, S., & Wedderburn, A. (1996). Maintaining Safety and High Performance on Shiftwork. *Applied Ergonomics*, *27*, 17-23.

Monk, T.H., & Kupfer, D.J. (2000). Circadian Rhythms in Healthy Aging-Effects Downstream from the Pacemaker. *Chronobiology International*, *17*, 55-68.

Nakano, T., Araki, K., Michimori, A., Inbe, H., Hagiwara, H., & Koyama, E. (2000). Temporal Order of Sleepiness, Performance and Physiological Indices during 19-h Sleep Deprivation. *Psychiatry and Clinical Neurosciences*, *54*, 280-290.

Patat, A., Rosenzweig, P., Enslen, M., Trocherie, S., Miget, N., Bozon, M.C., Allain, H., & Gandon, J.M. (2000). Effects of a new Slow Release Formulation of Caffeine on EEG, Psychomotor and Cognitive

Functions in Sleep-Deprived Subjects. *Human Psychopharmacology Clinical and Experimental*, *15*, 153-170.

Petz, B., & Vidacek, S. (1999). Studies of Psychophysiological and Temporal Conditions of Work. *Arhiv za higijenu rada i toksikologiju*, *50*, 405-421.

Phillipson, E.A. (1978). Respiratory Adaptations in Sleep. *Journal of Applied Physiology*, *40*, 895-902.

Phillipson, E.A., & Bowes, G. (1986). Control of Breathing during Sleep. In N.S. Cherniack & J.G. Widdicombe (Eds.), *Handbook of Physiology, Section III. The Respiratory System* (pp. 642-689). Bethesda, MD: American Physiological Society.

Piérard, C., Beaumont, M., Enslen, M., Chauffard, F., Tan, D.X., Reiter, R.J., Fontan, A., French, J., Coste, O., & Lagarde, D. (2001). Resynchronization of Hormonal Rhythms after an Eastbound Flight in Humans: Effects of Slow-Release Caffeine and Melatonin. *European Journal of Applied Physiology*, *85*, 144-150.

Pittendrigh, C.S. (1981). Circadian Systems: Entrainment. In J. Aschoff (Ed.), *Handbook of Behavioral Biology: Biological Rhythms* (pp. 95-124). New York, NY: Plenum Press.

Rechtschaffen, A., & Kales, A. (Eds.) (1968). *A Manual of Standardized Terminology. Techniques and Scoring System for Sleep Stages of Human Subjects*. Washington, DC: U.S. Government Printing Office.

Smolensky, M.H., Halberg, F., & Sargent, F., II. (1972). Chronobiology of the Life Sequence. In S. Itoh, K. Pogata, & H. Yoshimura (Eds.), *Advances in Climatic Physiology* (pp. 281-318). Tokyo: Igaku-Shoin.

Smolensky, M.H., Tatar, S.E., Bergmann, S.A., Losmann, J.G., Barnard, C.N., Dasco, C.C., & Kraft, I.A. (1976). Circadian Rhythmic Aspects of Human Cardiovascular Function: A Review by Chronobiologic Statistical Methods. *Chronobiologia*, *3*, 337-371.

Stadick, A., Bryans, R., Halberg, E., & Halberg, F. (1988). Circadian Cardiovascular Rhythms during Recumbency. In B. Tarquini & R. Vergassola (Eds.), *Social Diseases and Chronobiology* (pp. 191-200). Bologna: Esculapio.

Stepanova, S. (1986). *Biorhythmic Aspects of the Adaptation Problem*. Moscow: Nauka.

Stevens, R.G., & Rea, M.S. (2001). Light in the Built Environment: Potential Role of Circadian Disruption in Endocrine Disruption and Breast Cancer. *Cancer Causes Control*, *12*, 279-287.

Trinder, J., Whitworth, E., Kay, A., & Wilkin, P. (1992). Respiratory Instability during Sleep Onset. *Journal of Applied Physiology*, *73*, 2462-2469.

Voigt, E.D., Engel, P., & Klein, H. (1968). Über den tagesgang der körperlichen leistungfähigkeit. *Internationale Zeitschrift der Angewandte Physiologie*, *25*, 1-12.

Zulley, J. (2000). The Influence of Isolation on Psychological and Physiological Variables. *Aviation, Space, and Environmental Medicine*, *71* (Suppl 9), A44-A70.

3.2.2 Hydration

3.2.2.1 Definition and Measurement

Proper hydration is essential for optimal human performance. Euhydration refers to “normal” total body water (TBW), whereas hypohydration refers to a body water deficit. The term dehydration is used to refer to the dynamic process of body water loss (i.e., the transition from euhydration to hypohydration) (Greenleaf & Sargent, 1965; Sawka, 1992). The term hypovolemia defines a condition when blood volume is less than "normal".

3.2.2.2 Background

Adequate hydration is essential for maintaining fighting effectiveness, and several common operational stresses can result in relatively large alterations in TBW content and distribution. During most "normal" conditions, humans have little trouble maintaining optimal fluid balance. However, many factors such as sickness, physical exercise, climatic exposure (heat, cold, and altitude), and psychological strain can lead to significant disturbances in water balance (Sawka, 1988). Perhaps the best example involves heat stress and physical activity. For sedentary persons in temperate conditions, water requirements usually range from 2 to 4 L per day, and the kidneys primarily regulate fluid balance. For physically active persons who are exposed to heat stress, water requirements can often double (Sawka, Montain, & Latzka, 2001).

Water is the largest single constituent of the body (50 – 70% of body weight) and is essential for supporting the cardiovascular and thermoregulatory systems, and cellular homeostasis. TBW is distributed into intracellular fluid (ICF) and extracellular fluid (ECF) compartments. Exercise-heat stress not only stimulates fluid loss (primarily through sweating) but also induces electrolyte imbalances and renal function changes. As a result, fluid losses and gains with and without proportionate solute changes can occur. In addition, exercise-heat stress will alter transcompartmental and transcapillary forces that redistribute fluids between various compartments, organs, and tissues (Sawka et al., 2001). For these reasons, the accuracy of most methods used to assess hydration status is highly limited by the circumstances in which the measurements are made and the purposes for which they are intended.

TBW is the "gold standard" measurement to assess hydration status (Aloia, Vaswani, Flaster, & Ma, 1998; Lesser & Markofsky, 1979). TBW can be directly measured with doubly labeled water (DLW) and other dilution techniques. However, the requirement for expensive equipment and the associated technical problems make the use of these methods impractical. Although the choice of biomarker for assessing hydration status should ideally be sensitive and accurate enough to detect relatively small fluctuations in body water, the practicality of its use (time, cost, and technical expertise) is also of significant importance.

3.2.2.3 Effect on Performance

Both physical and cognitive performance are impaired proportionally to the magnitude of body water loss incurred (Sawka, 1988), but even small losses of body water (1 – 2% of body mass) have a measurable detrimental impact on physical work and negatively impact thermoregulation (Sawka, 1992; Sawka & Coyle, 1999).

3.2.2.4 Assessment Methods

Estimates of hydration are commonly made using (1) bioelectrical impedance analysis, (2) plasma indices, (3) urinalysis, and (4) changes in body weight. Given consideration to military field operational use, hydration assessment measurements are presented in order of increasing accessibility and practicality. Table 10 summarizes the advantages and disadvantages of each method.

Table 10: Biomarkers of Hydration Status

Measurement	Advantages	Disadvantages
Bioelectrical Impedance Analysis	Non-invasive Quick assessment	Measurements confounded by posture, diet, temperature, and fluid & electrolyte concentrations Invalid to assess hydration changes
Plasma Indices	Reliable for hyperosmotic dehydration and hyponatremia	Invasive measurement Moderately complex instrumentation
Urinalysis	Quick assessment Reliable measure 1st morning urine (U_{osm} and USG) Valid for screening of dehydration if combined with other indices	Unreliable for tracking acute changes in hydration Color influenced by diet, multivitamins, and medications
Body Weight	Quick assessment Simplest technique Easy to track in field exercises	Unreliable overtime due to changes in body composition (mass from lean body tissue)

3.2.2.4.1 Bioelectrical Impedance Analysis

Recently, bioelectric impedance analysis (BIA) has gained attention because it is simple to use and provides rapid, inexpensive, and non-invasive estimates of TBW (O'Brien, Young, & Sawka, 2002). Total body water volume is directly proportional to impedance (Berneis & Keller, 2000; Kushner, 1992; O'Brien et al., 2002). In practice, a small constant current is passed between electrodes spanning the body and the voltage drop between electrodes provides a measure of impedance (Kushner, 1992).

BIA does not have sufficient accuracy to validly assess moderate dehydration (~7% TBW) and loses resolution with isotonic fluid loss (O'Brien et al., 2002). In addition, since fluid and electrolyte concentrations can have independent effects on the BIA signal, the measurement can often provide grossly misleading values regarding hydration status (O'Brien et al., 2002). BIA has little application for the field assessment of hydration status.

3.2.2.4.2 Plasma Indices

Plasma volume decreases with dehydration; however, this response varies as a function of the type of dehydration (iso-osmotic or hyper-osmotic), physical activity, and the individual's heat acclimatization status and physical fitness (Sawka, 1988). Plasma volume changes can be estimated from hemoglobin and hematocrit changes; however, accurate measurement of these variables requires considerable controls for posture, arm position, skin temperature, and other factors (Sawka, 1988). If adequate controls are employed, plasma volume decreases proportionally with level of exercise-heat mediated dehydration.

In heat-acclimated persons undergoing exercise-heat mediated dehydration, resting plasma volume decreases in a linear manner that is proportional to the water deficit (Sawka & Coyle, 1999). These same levels will be maintained during subsequent physical exercise. If an iso-osmotic dehydration occurs, such as with altitude or cold exposure (O'Brien, Young, & Sawka, 1998; Sawka, 1992), then plasma osmolality changes will not follow TBW changes, and much larger plasma volume reductions will occur. The measurement of plasma osmolality and sodium requires phlebotomy (invasive), technical skill, and expensive instrumentation.

3.2.2.4.3 Urinalysis

Urinalysis is a frequently used clinical measure to distinguish between normal and pathological conditions. Urinary markers of hydration status include urine specific gravity (USG), urine osmolality (U_{Osmol}), and urine color. Urine specific gravity and osmolality are quantifiable and threshold values can provide some meaningful interpretation, whereas color is subjective and can be influenced by many factors including diet. It is important to recognize that the accuracy of these urinary indices in assessing chronic hydration status is improved when the first morning urine is used due to a more uniform volume and concentration (Sanford & Wells, 1962; Shirreffs & Maughan, 1998). Likewise, many factors such as diet, medications, exercise, climatic exposure, and timing can confound these indices.

The most widely used urine index is USG, measured against water as a standard (1.000 g/ml). Because urine is a solution of water and various other substances, normal values range from 1.010 to 1.030 (Armstrong et al., 1994; Popowski et al., 2001; Sanford & Wells, 1962). It has been suggested that a USG ≤ 1.020 represents a state of euhydration (Armstrong et al., 1994; Popowski et al., 2001). As a measure of chronic hydration status, USG appears to accurately reflect a hypohydrated state when in excess of 1.030 (Francesconi et al., 1987; Adolph, 1947; Armstrong et al., 1994; Popowski et al., 2001). However, considerable variability exists and no singular value can be used to determine a specific hydration level. U_{Osmol} also can provide an approximation of hydration status (Shirreffs & Maughan, 1998) since it is highly correlated with USG (Armstrong et al., 1994; Popowski et al., 2001), but the values are more variable.

3.2.2.4.4 Body Weight

Body weight (BW) measurements represent the simplest technique for rapid assessment of changes in hydration status. In our laboratory, we observe very small (< 1%) fluctuations in first morning BW when measured over consecutive days in young men taking food and fluid *ad libitum*. The stability of this measurement, coupled with the known losses of fluid that occur with exercise-heat exposure (primarily eccrine sweat), allows rapid changes in BW (incurred over hours) to be correctly attributed to water loss. Acute changes in BW are therefore a popular and reasonable field estimate of dehydration (Cheuvront, Haymes, & Sawka, 2002).

The level of dehydration is expressed as a percentage of starting body weight [(ΔBW/startBW) × 100] rather than as a percentage of total body water (TBW) since TBW ranges from 50 – 70% of body weight. This technique assumes that (1) starting BW represents a euhydrated state, and (2) 1ml of sweat loss represents a 1g change in weight (i.e., the specific gravity of sweat is 1.000 g/ml). As an acute measure, first morning BW is still limited by changes in bowel habits. BW is also limited as a tool for long-term assessment of hydration status since changes in body composition (fat and lean mass) that occur with chronic energy imbalance are also reflected grossly as changes in BW. Clearly, the use of daily body weight should be used in combination with another hydration assessment technique (first morning urine) to dissociate gross tissue losses from water losses if long-term hydration status is of interest.

3.2.2.5 Practical Applications

Under most conditions, day-to-day body mass changes (<2%) and first morning urine specific gravity (<1.030) when used together provide an approximate indication that an individual is dehydrated (see Table 10). However, plasma osmolality changes can provide more reliable information regarding hydration when greater precision is required. Moreover, BIA should not be used to assess hydration status in the field for reasons previously described. It is possible that other technological advances may allow evaluation of other measures (e.g., muscle water content) that hold promise as hydration indices.

The views, opinions, and/or findings contained in this publication are those of the authors and should not be constructed as an official Department of the Army position, policy, or decision unless so designated by other documentation.

3.2.2.6 References

Adolph, E.F. (1947). *Physiology of Man in the Desert*. New York: Intersciences, Inc.

Aloia, J.F., Vaswani, A., Flaster, E., & Ma, R. (1998). Relationship of Body Water Compartments to Age, Race, and Fat-Free Mass. *Journal of Laboratory and Clinical Medicine*, *132*, 483-490.

Armstrong, L.E., Maresh, C.M., Castellani, J.W., Bergeron, M.F., Kenefick, R.W., LaGasse, K.E., & Riebe, D. (1994). Urinary Indices of Hydration Status. *International Journal of Sport Nutrition*, *4*, 265-279.

Berneis, K., & Keller, U. (2000). Bioelectrical Impedance Analysis during Acute Changes of Extracellular Osmolality in Man. *Clinical Nutrition*, *19*, 361-366.

Cheuvront, S.N., Haymes, E.M., & Sawka, M.N. (2002). Comparison of Sweat Loss Estimates for Women during Prolonged High-Intensity Running. *Medicine and Science in Sports and Exercise*, *34*, 1344-1350.

Francesconi, R.P., Hubbard, R.W., Szlyk, P.C., Schnakenberg, D., Carlson, D., Leva, N., Sils, I., Hubbard, L., Pease, V., & Young, J. (1987). Urinary and Hematologic Indexes of Hypohydration. *Journal of Applied Physiology*, *62*, 1271-1276.

Greenleaf, J.E., & Sargent, F., 2nd. (1965). Voluntary Dehydration in Man. *Journal of Applied Physiology*, *20*, 719-724.

Kushner, R.F. (1992). Bioelectrical Impedance Analysis: A Review of Principles and Applications. *Journal of the American College of Nutrition*, *11*, 199-209.

Lesser, G.T., & Markofsky, J. (1979). Body Water Compartments with Human Aging using Fat-Free Mass as the Reference Standard. *American Journal of Physiology*, *236*, R215-220.

O'Brien, C., Young, A.J., & Sawka, M.N. (1998). Hypohydration and Thermoregulation in Cold Air. *Journal of Applied Physiology*, *84*, 185-189.

O'Brien, C., Young, A.J., & Sawka, M.N. (2002). Bioelectrical Impedance to Estimate Changes in Hydration Status. *International Journal of Sport Medicine*, *23*, 361-366.

Popowski, L.A., Oppliger, R.A., Patrick Lambert, G., Johnson, R.F., Kim Johnson, A., & Gisolf, C.V. (2001). Blood and Urinary Measures of Hydration Status during Progressive Acute Dehydration. *Medicine and Science in Sports and Exercise*, *33*, 747-753.

Sanford, S.A., & Wells, B.B. (1962). *The Urine* (Vol. 13). Philadelphia: W.B. Saunders, Co.

Sawka, M.N. (1988). Body Fluid Responses and Hypohydration during Exercise-Heat Stress. In K.B. Pandolf, M.N. Sawka & R.R. Gonzalez (Eds.), *Human Performance Physiology and Environmental Medicine at Terrestrial Extremes* (pp. 227-266). Indianapolis, IN: Benchmark Press, Inc.

Sawka, M.N. (1992). Physiological Consequences of Hypohydration: Exercise Performance and Thermoregulation. *Medicine and Science in Sports and Exercise*, *24*, 657-670.

Sawka, M.N., & Coyle, E.F. (1999). Influence of Body Water and Blood Volume on Thermoregulation and Exercise Performance in the Heat. *Exercise and Sport Science Reviews*, *27*, 167-218.

Sawka, M.N., Montain, S.J., & Latzka, W.A. (2001). Hydration Effects on Thermoregulation and Performance in the Heat. *Comparative Biochemistry and Physiology: Molecular Integrative Physiology*, *128*, 679-690.

Shirreffs, S.M., & Maughan, R.J. (1998). Urine Osmolality and Conductivity as Indices of Hydration Status in Athletes in the Heat. *Medicine and Science in Sports and Exercise*, *30*, 1598-1602.

3.2.3 Illness

3.2.3.1 Definitions

By their very nature, illnesses produce performance decrements in afflicted individuals. While severe symptoms may lead to bed rest and even hospitalization, individuals with less severe symptoms may continue to work. Besides the risk of spreading their illness to co-workers, ill workers may not perform their jobs at the levels realized when they are healthy. This not only lowers their efficiency, but, in some cases, may also result in errors that put themselves, co-workers, and the systems they operate in jeopardy. Severe chronic illnesses are usually not of concern to day-to-day operations because these individuals are aware of their conditions and do not work beyond their capabilities. Of concern are rapid onset illnesses such as heart attacks that may occur at the work place, and more transient illnesses such as colds, whose symptoms may be judged by the ill person to not be serious enough to warrant missing work. In operational situations, it is possible that the performance consequences of these minor illnesses may be judged inconsequential and the ill individual is expected to work. Closer examination of some of these minor illnesses shows that performance may be affected.

It has been estimated that 10 to 12 percent of all work absences are due to colds and influenza (Smith, 1992). It is well known that many individuals with colds still report to work. The motivation behind working when ill varies, but, nevertheless, working while ill can be associated with sub-par work performance. This performance reduction may also result in accidents and errors that can affect others. The onset of illness may prompt individuals to seek relief from symptoms through the use of over-the-counter (or prescription) medications, which often produce side effects that also degrade performance. The job environment itself may induce transient illness in some individuals. For example, debilitating motion sickness, simulator-induced vertigo, and virtual reality-induced vertigo may occur in healthy individuals and may be induced by the work environment (Bullinger, Bauer, & Braun, 1997; Griffin, 1997). Other illnesses of interest brought about by occupational activities include dehydration and acute mountain sickness.

Because the level of performance degradation and the particular cognitive modality affected varies, it is difficult to construct general guidelines concerning whether or not a given individual should work when ill. Probably the most widespread of the minor illnesses that have been shown to produce performance effects are the common cold and influenza. Not only do the symptoms of these two illnesses differ, but their effects on performance also differ.

3.2.3.2 Background

Ill individuals typically suffer performance decrements. The effects of these decrements can vary from minor to catastrophic. Furthermore, in the case of major illnesses, the performance of co-workers can be influenced. For example, a severe cardiac episode can disrupt normal procedures and lead to behavior that is disruptive of co-worker performance. This can be especially serious in certain jobs that involve the well being of others such as aircraft crews. Ill pilots are not only incapacitated so that they are not able to perform their duties but their illness distracts other crew members from properly performing their duties, which can lead to serious consequences (Baker, 1999).

3.2.3.2.1 Chronic vs. Acute Illness

Common colds and influenza are much more prevalent in the work place than are catastrophic illnesses. As such, they probably represent the most widespread illnesses with which operators still report to work

(Smith, 1992). Other illnesses that have effects on performance during their acute and chronic phases have been studied. Egan and Goodwin (1992) described the effects of HIV and AIDS. Hall and Smith (1996a) described the differential performance effects of acute and chronic infectious mononucleosis. There was a general increase in the acute infected subjects' negative affect in a task in which subjects did not know prior to stimulus presentation where to respond, and their performance was worse than that of controls. This effect was similar to the profile reported in influenza patients. Chronic infectious mononucleosis was associated with memory problems that were similar but not the same as have been reported in cold sufferers. It is noteworthy that the strength of the performance decrements was not highly correlated with the level of symptom severity (Hall & Smith, 1996; Matthews, Warm, Dember, Mizoguchi, & Smith, 2001; Smith, Tyrrell, Coyle, & Willman, 1987). Cold and influenza infections seem to produce specific performance and subjective effects rather than generalized deficiencies. Colds and influenza are associated with different patterns of cognitive processing and subjective feelings. This effect is in agreement with reports showing that other environmental stressors result in differential effects rather than global effects common to all stressors (Hockey & Hamilton, 1983; Hockey, 1986). For practical purposes, one must determine the exact nature of the illness in order to determine the nature of potential performance deficits when making decisions about readiness to perform.

3.2.3.2.2 Minor Illnesses

Motion sickness is a condition of dizziness (vertigo), nausea, and vomiting brought about by movement. Motion sickness may be caused by the actual work milieu such as aircraft and ship motion or by simulator and virtual environments. The condition is believed to occur because of disparate input to the brain from movement-sensing organs: the eyes, inner ears, skin, and muscle. Thus, motion sickness is most commonly observed in a vehicle wherein an individual may have limited exposure to the outside visual scene. The eyes may indicate a stable world, while the vestibular system registers movement. The brain receives incongruent sensory information, and a feeling of discomfort and dizziness may emerge.

Dehydration is characterized by a reduction in body fluids to the extent that normal bodily functions are impacted and performance is compromised. Loss of fluids may result from inadequate intake or loss due to vomiting, diarrhea, excessive sweating, or excessive urine output (polyuria). Typically associated with other maladies (e.g., heat stress, influenza), the symptoms associated with dehydration sickness include dry and sticky mucus membranes of the mouth, malaise, and intense thirst. More serious conditions are associated with a sunken appearance to the eyes and deterioration of skin elasticity. Seizures may occur with prolonged dehydration. Physiological changes associated with dehydration include low blood pressure and rapid heart rate. Analysis of blood serum for electrolyte levels and urine for various markers (e.g., specific gravity) will indicate dehydration with a high degree of accuracy. Unless dehydration is the result of a medical condition such as diabetes, it can usually be linked to specific environmental conditions and behaviors. Lack of proper hydration in the face of exertion is a typical cause of dehydration.

Acute Mountain Sickness (AMS) may occur in individuals who ascend to over 2500 meters (8000 feet) without physiologically acclimatizing to higher altitude. This illness occurs as a function of prolonged hypoxia as the body attempts to maintain normal blood oxygen saturation levels at higher altitudes. AMS is often exacerbated by physical exertion and inadequate hydration, both of which occur readily at altitude. The onset of AMS is associated with headache and nausea/vomiting, severe fatigue, dizziness, confusion and/or staggering gait. Individuals may become incapacitated and require assistance descending to lower altitudes. The life-threatening conditions of High Altitude Pulmonary Edema (HAPE) and High Altitude Cerebral Edema (HACE) typically follow untreated AMS.

3.2.3.2.3 Chronic Illness

Some diseases, such as coronary artery disease, are seen as such potential risks that operators with the disease may be excluded from certain jobs. Cardiovascular disease is the leading cause of disqualification of aviators worldwide (Smalley, Loecker, Collins, Prince, and Browning, 2000). The potential level of incapacitation that could result from a heart attack is so high that potential operators with the disease are excluded from being selected. This decision is especially pertinent in commercial transportation where the lives of others are at stake. Pilots are required to obtain periodic physical examinations to maintain their flying status.

3.2.3.3 Effects on Performance

Depending on the type of illness and its severity, the performance effects range from minor to disastrous. Illnesses that cause total incapacitation of the operator of a single-operator system result in cessation of performance. Serious illness in one member of a multi-person crew typically means that other crew members must assume the duties of the ill operator. Furthermore, the illness of one member of a crew can cause other crew members to become distracted as they try to assist the ill operator. This reduces the effectiveness of the crew members who are trying to help the ill colleague.

3.2.3.3.1 Acute Illness

Much of the available research on the performance effects of illness that is relevant to the present discussion has involved the study of the common cold and influenza. Both of these diseases are quite common and also ones in which operators may continue to work at their jobs even though they are ill. A.P. Smith has conducted much of the relevant research over the past 14 years (Smith, Thomas, & Whitney, 2000). Smith and co-workers have studied both experimentally induced and naturally occurring colds and influenza. They report that these illnesses influence different aspects of the ill person's mood and performance on laboratory tasks.

Colds affected attentional tasks and psychomotor functioning as demonstrated by poorer tracking performance and reaction times but not working or semantic memory (Hall & Smith, 1996b). The reaction time effects have been demonstrated with both simple and choice reaction time tasks. These effects may be due to the low arousal states associated with the colds because the effects could be reversed with the use of caffeine (Smith, Thomas, Perry, & Whitney, 1997). Reduced accuracy during a signal detection task and increased subjective workload were reported in subjects with colds by Matthews et al. (2001). They felt that ill operators were especially vulnerable to errors on components of tasks that require attention. Subjects with colds also score lower on alertness and sociability scales and higher on a tenseness scale (Smith et al., 2000).

Alcohol ingestion was reported to produce different effects on healthy vs. cold-suffering subjects (Smith, Whitney, Thomas, Brockman, & Perry, 1995). Alcohol ingestion improved the subjective mood of healthy subjects but resulted in increased negative moods in subjects with a cold. The alcohol produced faster but less accurate responses when the subjects with a cold performed a sustained attention task. Alcohol also impaired performance on a focused attention task. This suggests that cold sufferers should be wary when offered alcohol or other drugs to improve their symptoms. Although drug effects are not within the scope of this section, many cold medications produce an increased likelihood of drowsiness and dizziness. Matthews et al. (2001) point out that subjective measures may not yield an accurate picture of the degree of performance impairment. They suggest that direct measures of simple reaction time are appropriate in jobs where psychomotor performance is critical.

AMS is associated with headache and nausea/vomiting, severe fatigue, dizziness, confusion and/or staggering gait. The effects of AMS on performance range from mild to profound. AMS is usually avoided by carefully acclimatizing to altitude. Common practice is to ascend no more than 300 meters (1000 feet) per day at altitudes of 3000 meters (10,000 feet) or greater. Other regimes call for climbing much higher but descending to an intermediate altitude to sleep. Immediate descent to lower altitude is necessary if symptoms of HACE (inability to reason, confusion, loss of coordination) or HAPE (chest congestion, cough, breathless at rest) are present, or if AMS symptoms worsen. Further ascent in the face of any AMS symptomology is suicidal. Acetazolamide (Diamox) is a sulfonamide medication that forces bicarbonate excretion from the kidneys and works to acidify the bloodstream. The acidification is believed to stimulate respiration and accelerate acclimatization. Prophylactic use of Diamox among hikers and climbers is fairly commonplace, but not without some risky side effects or possible allergic reaction to the sulfa-based compound.

3.2.3.3.2 Interactions with Other Factors

The interaction of illness with other factors that are known to produce degraded performance has also been reported. Smith et al. (2000) studied performance on laboratory tasks in subjects at the beginning of the work day, after lunch, and at the end of the work day. Overall, they found that cold sufferers showed slowed reaction times, decreased alertness, and decreased sociability over the course of the day. These results again demonstrate the interaction of cold effects with other stressors. This suggests that operators who may be borderline at the beginning of the work shift may exhibit degraded performance later on. This effect is especially noteworthy for operators who must perform at high levels during their entire shift.

Vertigo-related nausea and discomfort can interfere with operator performance and, in some cases, be totally debilitating. The incidence of motion sickness varies among crew members and may be related to control over the situation and physical location in the vehicle. Motion sickness in pilot trainees can result in removal from training or delays in training. Several programs have proven successful in reducing motion sickness (Turner, Griffin, & Holland, 2000). Some individuals may be more prone to motion sickness, and some people may experience the symptoms days after exposure to a significant motion stimulus. Indeed, “simulator sickness,” a motion sickness known to afflict aviators in flight simulator training, can occur for some period after the training regardless of whether the simulator is motion-capable. Effects on performance are dependent upon the severity of sickness. Vigilance, problem solving, and the ability to carry out physical tasks are all impacted. Motion sickness can be treated with the use of antihistamines or a scopolamine patch. These treatments are most effective when administered before exposure to motion.

Influenza sufferers displayed impaired stimulus detection when the stimuli appeared at unpredictable times and places, but influenza did not impact psychomotor task performance. Influenza decreased both the speed and accuracy of a stimulus detection task. A 57% increase in reaction time was found, which is five to ten times the magnitude of previously reported effects of moderate alcohol intake (Smith et al., 1987). Smith et al. (2000) suggested that colds might produce greater performance deficiencies in operators who are already prone to distraction and who are not highly motivated. These operators are already at risk for increased errors, and the presence of colds or influenza may magnify the risk, resulting in a higher probability of errors.

3.2.3.4 Assessment Methods

Much of the research on the effects of colds and influenza has been carried out using laboratory tasks. Psychomotor tracking, simple and complex reaction time, variable fore-period reaction time, signal detection, and pursuit tracking tasks have been used. Focused attention, categorical search, vigilance, and numerous memory tasks have also been used. Various subjective mood scales have been used to assess the effects of colds and influenza. See the various articles cited in the reference section for more details on the various tests that have been used.

Real-time assessment of fluid volume and electrolyte concentrations for detection of dehydration onset is currently not possible. Assessment for tachycardia is reasonable, but would require assessment of activity level to determine a mismatch between exertion level and cardiac effort. Similarly, periodic assessment of blood pressure for hypotension would also require some control for activity level. The most straightforward way to assess the functional state of an operator at risk for dehydration may be from the combined measures of fluid intake (volume from a wearable re-hydration system), exertion (via accelerometer and/or heart rate measures), and environmental conditions. Such an approach is applicable for physical work in both confined spaces or in remote field settings.

One measurable indicator of the vertigo associated with motion sickness is nystagmus, a rhythmic shifting of the eyes from side to side in response to head motion. Clinically, physicians will look for nystagmus, and thus susceptibility to vertigo, by moving a patient's head quickly and then inspecting the eyes. Eye activity monitoring systems would be suitable for the detection of nystagmus, and thus might be useful for indicating the onset of vertigo. Seated systems operators within moving vehicular platforms such as ships and tanks would be most susceptible to motion sickness, and an eye activity-based vertigo detection system might be useful in those domains. At this writing, however, such an application of eye tracking technology has not been implemented.

A monitoring system capable of detecting the onset of AMS might be possible in the near future. Non-invasive blood oxygenation measurement, combined with cardiac output and possibly respiration rate measures, might form the basis for such a system. Currently, such measures may not provide ample sensitivity to detect subtle hypoxia as a function of effort at high altitude, but these technologies continue to improve. Secondary to AMS, a method of monitoring for HAPE, which often occurs during sleep, would be useful. Changes in breathing patterns and indicators of chest congestion would be useful in this regard. A system dedicated to indicating the *likelihood* of AMS onset might be a more realistic near-term possibility. Using a Global Positioning System capability to monitor distance and elevation changes, a computer with fundamental information about an individual (weight with pack, fitness level, terrain type, recent extent of time at altitude) could readily estimate caloric expenditure over time. Combined with saturated oxygen, respiration rate, and heart rate data, such a system could conceivably alert an individual to AMS risk. As with any system of this type, accuracy would be greatly enhanced if previous incidents of actual AMS onset could be logged by the system and matched to various measurement parameters.

3.2.3.5 References

Baker, S.M. (1999). Human Factors and Accident Causation: The Interface between Illness and Performance. *European Heart Journal Supplements*, *1 (Suppl. D)*, D15-D18.

Bullinger, H., Bauer, W., & Braun, M. (1997). Virtual Environments. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1725-1759). New York: John Wiley.

Griffin, M.J. (1997). Vibration and Motion. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 828-857). New York: John Wiley.

Hall, S.R., & Smith, A.P. (1996a). Behavioural Effects of Infectious Mononucleosis. *Neuropsychobiology*, *33*, 202-209.

Hall, S.R., & Smith, A.P. (1996b). Investigation of the Effects and After-Effects of Naturally Occurring Upper Respiratory Tract Illnesses on Mood and Performance. *Physiology and Behavior*, *59*, 569-577.

Hockey, G.R.J. (1986). A State Control Theory of Adaptation to Stress and Individual Differences in Stress Management. In G.R.J. Hockey, A.W.K. Gaillard, & M.G.H. Coles (Eds.), *Energetics and Human Information Processing* (pp. 285-298). Dordrecht, The Netherlands: Martinus Nijhoff.

Hockey, G.R.J., & Hamilton, P. (1983). The Cognitive Patterning of Stress States. In G.R.J. Hockey (Ed.), *Stress and Fatigue in Human Performance* (pp. 331-362). Chichester, England: Wiley.

Matthews, G., Warm, J.W., Dember, W.N., Mizoguchi, H., & Smith, A.P. (2001). The Common Cold Impairs Visual Attention, Psychomotor Performance and Task Engagement. In *Proceedings of the Human Factors and Ergonomics Society* (pp. 1377-1381). Santa Monica, CA: Human Factors and Ergonomics Society.

Smalley, B., Loecker, T.H., Collins, T.R., Prince, T.S., & Browning, S.R. (2000). Positive Predictive Value of Cardiac Fluoroscopy in Asymptomatic U.S. Army Aviators. *Aviation, Space, and Environmental Medicine*, *71*, 1197-1201.

Smith, A.P. (1992). Colds, Influenza and Performance. In A.P. Smith & D.M. Jones (Eds.), *Handbook of Human Performance: Vol. 2: Health and Performance* (pp. 197-218). London: Academic.

Smith, A., Thomas, M., & Whitney, H. (2000). Effects of Upper Respiratory Tract Illnesses on Mood and Performance over the Working Day. *Ergonomics*, *43*, 752-763.

Smith, A.P., Tyrrell, D.A.J., Coyle, K., & Willman, J.W. (1987). Selective Effects of Minor Illnesses on Human Performance. *British Journal of Psychology*, *78*, 183-188.

Smith, A.P., Whitney, H., Thomas, M., Brockman, P., & Perry, K. (1995). A Comparison of the Acute Effects of a Low Dose of Alcohol on Mood and Performance of Healthy Volunteers and Subjects with Upper Respiratory Tract Illnesses. *Journal of Psychopharmacology*, *9*, 267-272.

Turner, M., Griffin, M.J., & Holland, I. (2000). Airsickness and Aircraft Motion during Short-Haul Flights. *Aviation, Space, and Environmental Medicine*, *71*, 1181-1189.

3.2.4 Mental Fatigue

3.2.4.1 Definitions and Measurement

Research on mental fatigue has suffered from a lack of conceptual clarity and problem specification. There have been long-standing problems of definition, and few systematic attempts to develop a satisfactory theoretical foundation. More so than with most psychological constructs, there has never been agreement about what fatigue is, except that it is characterized by a subjective state of tiredness. Early in the twentieth century, Muscio (1921) suggested that the term “fatigue” be completely abandoned, such were the difficulties of finding an adequate basis for definition.

The most widely adopted operational definition of mental fatigue is that of a temporary decrement in ongoing task activity, generally occurring over a relatively brief period (up to a few hours). It is generally agreed that mental fatigue refers specifically to *a state of tiredness that develops over time, especially when a person has been carrying out a mental task or dealing with stressful events*. This distinguishes it from long-term conditions such as chronic fatigue and burnout, and from other kinds of short-term fatigue. In addition to the marker of subjective tiredness, a second defining feature of mental fatigue is usually taken to be the observation of a decrement in task performance. However, this is not always reliable, because of the use of performance protection strategies by operators (Hockey, 1997), especially in highly skilled work. More diagnostic of fatigue is the detection of a *fatigue after-effect* – a reduction in task engagement following work, characterized by a preference for low effort strategies (Holding, 1983).

3.2.4.2 Background

Fatigue is a pervasive state of the human condition, found in many different forms. Although all forms are characterized by a feeling of tiredness, distinctions are normally made between fatigue that comes from doing mental tasks, from physical work, or through sleep disturbances. Yet, the various effects may have much in common. For example, it is thought that the limiting condition for tolerance of physical fatigue is not muscular strain but a loss of cognitive control (see Holding, 1983). Considering mental fatigue, there are reasons for making further distinctions according to the nature of the prevailing demands. Several different forms of fatigue are recognized in human factors research on performance degradation, and are often confused in the discussion of applied issues. Fatigue associated with sustained task performance and cognitive demands is usually referred to as mental (or cognitive) fatigue. We can go further and ask whether the same kind of fatigue process is activated as a response to different kinds of tasks – on the one hand, heavy demands over short periods (high workload decrements); on the other, light demands over very long periods (vigilance-type decrements).

In any case, work-based fatigue is usually distinguished from fatigue defined operationally by sleep disturbances, or from natural variation in sleep schedules. It may also be necessary to distinguish fatigue associated with sleep deprivation from fatigue based on diurnal rhythms (e.g., Folkard & Åkerstedt, 1989). Of course, to further complicate matters, sleep disruptions may also be caused by both physical and mental work (of both kinds). We may also distinguish emotional fatigue associated specifically with emotional overload and *burnout* (Cherniss, 1980). Because of the links between these different states, any research on mental fatigue must recognize the possible influence of these other sources of fatigue. A simple representation of these relationships is shown in Figure 3. Each of the three stressors gives rise to a specific form of fatigue, with separate feedback loops indicating the corrective action necessary to reduce that kind of fatigue: for mental fatigue, rest (from tasks) or a change of task; for sleepiness, sleep; and for physical fatigue, physical rest. It is likely that all three stressors contribute to what is experienced as a generalized fatigue state, and may even give rise to a final common path (FCP) that determines the general effects on OFS and performance.

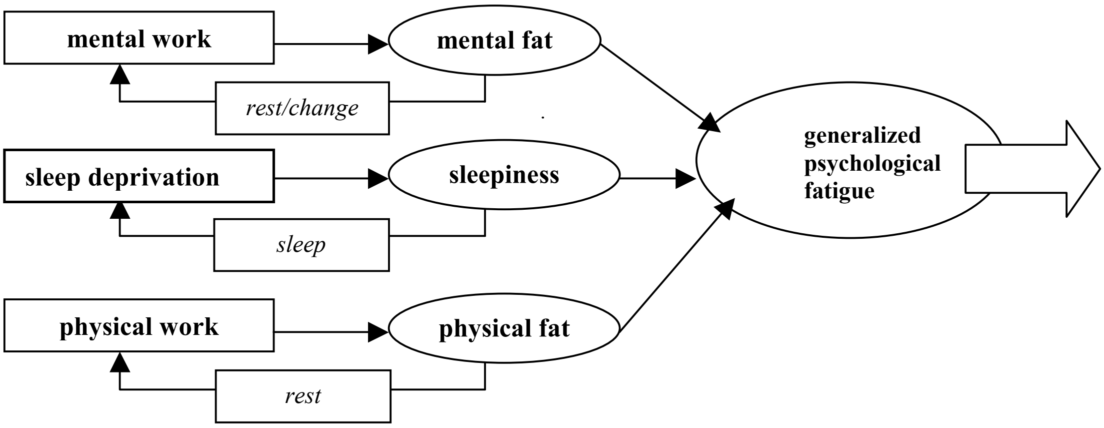

Figure 3: Possible Relationship between Different Sources of Fatigue.

In the context of modern work, mental and sleep-based fatigue are the most relevant problems. Many industrial and military operations are likely to give rise to both kinds of state, although little attempt has been made in relevant applied work settings to assess either their separate effects or the nature of their interaction. Mental and sleep-based fatigue conditions clearly have their origins in different kinds of environmental constraints, although there are strong functional links between the work/rest and sleep/wake cycles. It is therefore important, from both a practical and theoretical viewpoint, to consider their joint effects. Are the effects of prolonged work periods greater, for example, in individuals who have not slept for one or two nights, or whose sleep has been disrupted systematically over a long period? Is the impairment observed during a new period of night-shift work greater after a period of heavy cognitive demands? There is no direct evidence to address these questions, and in practical operational contexts (e.g., Fröberg, Karlsson, Levi, & Lidberg, 1975; Haslam, 1982), such interactions do not appear to have been studied.

3.2.4.3 Effects on Performance

Mental fatigue is generally associated with performance decrements, but the association is often drawn loosely. In its typical effect on loss of engagement with tasks, and impaired performance, mental fatigue has a superficial similarity to states such as boredom, distraction, and loss of task motivation. In earlier analyses of decrement (Welford, 1965), the effects of boredom were attributed to information underload (too low a level of relevant task inputs, as in vigilance) and the effects of fatigue were attributed to overload (high event rates, multiple tasks). This distinction has not been adequately tested, and the boundary conditions are unclear. For example, sustained attention on a low event rate monitoring task demands a high level of concentration and maintenance of attention, which can give rise to mental fatigue, especially where performance is maintained by a high level of effort. Likewise, a high-load task in which an individual does not engage can be boring and lead to decrement through inattention. Nevertheless, it is likely that fatigue effects will be more probable in overload situations because of the difficulty of keeping up with the high level of demands with few opportunities for rest (see Hockey, 1986).

Following earlier analyses of skill decrement by Bartlett (1953) and Broadbent (1958), the performance effects of fatigue are no longer confined to observations of reductions in output or general slowing of performance. Instead, researchers have looked for more subtle effects, reflected in the patterning or timing of responses. Some of the earliest work on reaction time (RT) identified fatigue effects in prolonged color naming (Bills, 1931). The effect was not an increase in mean RT over time, but an increase in the number of extremely slow responses (called ‘blocks’). This effect has been replicated in an extensive series of studies using the 5-choice serial reaction task (see Broadbent, 1971), where long responses are referred to as “gaps,” and in sleep deprivation studies (e.g., Williams, Lubin, & Goodnow, 1959), where they are called “lapses.” Such slowing can be detected for several seconds before the event, and may be part of an adaptive strategy for resetting attention control (Bertleson & Joffe, 1963; Rabbitt, 1981).

Another branch of early work involving skilled performance demonstrated fatigue effects as changes in the pattern of performance on different components of complex tasks. Some of the earliest effects of this kind were demonstrated on a simple cockpit simulator, and took the form of a neglect of peripheral (or less important) instruments with prolonged work (Bartlett, 1943; Davis, 1948). Similar results (sometimes called “narrowing of attention”) have been found in other multi-component tasks (Hockey, 1986). Bartlett’s analysis of mental fatigue remains central to the “explanation” of degradation in complex work and high-workload tasks, especially where the task session is prolonged and unbroken by rests. Sometimes, more specific explanatory concepts are preferred, particularly where workload is not excessive. For example, decrements in vigilance tasks are not normally attributed to fatigue, although some reviews of fatigue include vigilance because of the accompanying requirement for prolonged work (Craig & Cooper, 1992; Holding, 1983). One of the problems with the within-task decrement definition is that there is normally no independent assessment of the presence of a fatigue state. In this usage the term has no explanatory value – equivalent to using arousal or effort to “explain” unexpected improvements in performance. This problem was recognized by Muscio (1921), and restated by Broadbent (1979).

One of the strongest sources of evidence about the nature of the fatigue state is the demonstration that it is associated with resistance to (further) effort on tasks carried out after the operational work period has ended. This was first pointed out by Thorndike (1900), and frequently emphasized by other reviewers (e.g., Bartley & Chute, 1947), although it has only become well established through more recent analyses such as those of Holding and colleagues (see Holding, 1983). For example, given a choice of options, fatigued individuals adopt less effortful strategies to solve a problem, and seek less information before making a judgment (Webster, Richter, & Kruglanski, 1996). A modern perspective on mental fatigue is that resistance to effort (on post-work activities) is the main defining feature of the state. A major series of studies carried out in the 1950s (Chiles, 1955) found no reliable effects on pursuit tracking from fatigue induced by up to 2 days continuous work on a flight simulator. The explanation for this result appears to be that subjects were able to overcome their fatigue state momentarily by additional effort. Unless specific measurements are made, it is not possible to detect this compensatory activity. However, as Holding (1983) showed, when subjects are presented with a choice of two equally acceptable ways of responding, they are more likely to select the less-demanding option when tired.

3.2.4.4 Assessment Methods

Not surprisingly, given the lack of agreement about the status and definition of mental fatigue, it cannot be assessed directly, except where it is defined solely through subjective reports. Indirect markers of fatigue can sometimes be inferred from assessment of task performance and psychophysiological measures.

3.2.4.4.1 Subjective Reports

Since fatigue is essentially a feeling of tiredness, subjective reports provide the most direct index of the level of fatigue. Surprisingly, there is no recognized standardized test, most studies relying on ad hoc research questionnaires administered at the end of the work period, asking participants to indicate "how tired" or "how fatigued" they are. Either a Likert scale or visual analogue scale (VAS) is appropriate for this assessment, although reliability is improved by the use of several items rather than just one. It is usual to use bipolar scales, with items labeled "lively" or "energetic" at one end and "tired" or "fatigued" at the other end (Hockey, 1996; Pearson, 1957), although direct comparisons with monopolar scales have not been made.

3.2.4.4.2 Task Probes (After Effects)

It is clear from the above discussion that task measures during the work session do not provide reliable assessment of fatigue. To have true explanatory value, it is necessary to show that fatigue induction produces effects that extend beyond the situation in which the fatigue develops. However, while the demonstration of after-effects remains the acid test for the effects of mental fatigue, the two kinds of decrement are part of the same process of adaptation to work demands. They are linked through an understanding of the dynamics of the individual's response to the task environment, as driven by his or her own set of goals (Hockey, 1997).

Essentially the same processes are involved whether in response to work demands or stressors, or whether the goals are externally or internally driven. In either case, mental fatigue results from the (effortful) requirement to maintain orientation towards tasks or goals under changing environmental constraints. After-effects are assessed by carrying out further tasks in the period immediately following the main work session. This approach has been used by a number of researchers (Hockey & Wiethoff, 1989; Jongman, Meijman, & De Jong, 1999; Webster et al., 1996). Again, there are no standard tests for after-effects, although the general principles for designing such tests are becoming clearer. The tests should present the subject with a choice of two plausible ways of doing the task – a safe method that makes high demands on effort, and a more risky but low-effort alternative. Hockey and Wiethoff developed a drug prescription task for testing workload effects in doctors. When the doctors were unsure about whether medication was correct they could check back to a screen showing drug administration details (slow but safe) or simply guess (fast but risky). In practice, doctors did not guess, but more tired doctors checked less often and spent less time checking. Recent work by Meijman and colleagues (Meijman, Mulder, van Dormolen, & Cremer, 1992) has shown marked after-effects of demanding mental work on a fault diagnosis task. Induced fatigue decreased the use of an elaborate but safe hypothesis-testing strategy, in favor of a less effective strategy involving guessing. Jongman et al. (1999) suggested that fatigue involves a loss of the control processes that maintain activation of tasks in working memory, triggering compensatory changes in information processing strategy.

3.2.4.4.3 Psychophysiological Markers

There are no unambiguous markers of mental fatigue, although a number of measures have been assumed to be indicative of such a state. One problem is that it is difficult to distinguish the effects of stress from fatigue, so that many stress measures are also often used to infer fatigue (notably sympathetic activation). More reliably, increased levels of urinary or salivary cortisol, and sometime urinary adrenaline have been found to increase with fatigue associated with high workload tasks (Frankenhaeuser & Johansson, 1981), demanding workdays (Meijman, 1997), or working in industrial noise (Melamed & Bruhis, 1996). As yet, there appears to be no clear EEG marker. Measures of increased relative power in low-frequency bands may be indicative of low alertness. However, this is typically a passive state, and is unlikely to be the same as mental fatigue, which represents a reactive disengagement from task goals. Finding appropriate EEG indices of fatigue appears to be a major requirement for effective work on fatigue in operational conditions.

3.2.4.5 References

Bartlett, F.C. (1943). Fatigue following Highly Skilled Work. *Proceedings of the Royal Society (B)*, *131*, 147-257.

Bartlett, F.C. (1953). Psychological Criteria of Fatigue. In W.F. Floyd & A.T. Welford (Eds.), *Symposium on Fatigue*. London: Lewis.

Bartley, S.H., & Chute, E. (1947). *Fatigue and Impairment in Man*. New York: McGraw Hill.

Bertleson, P., & Joffe, R. (1963). Blocks in Prolonged Serial Responding. *Ergonomics*, *6*, 109-116.

Bills, A.G. (1931). Blocking: A New Principle in Mental Fatigue. *Journal of Psychology*, *43*, 230-245.

Broadbent, D.E. (1958). *Perception and Communication*. London: Pergammon.

Broadbent, D.E. (1971). *Decision and Stress*. London: Academic Press.

Broadbent, D.E. (1979). Is a Fatigue Test Now Possible? *Ergonomics*, *12*, 1277-1290.

Cherniss, C. (1980). *Staff Burnout: Job Stress in the Human Services*. Beverly Hills: Sage.

Chiles, W.D. (1955). Experimental Studies of Prolonged Wakefulness. *WADC Technical Report*, No. 55-395. Dayton, OH: Wright Air Development Center.

Craig, A., & Cooper, R.E. (1992). Symptoms of Acute and Chronic Fatigue. In A.P. Smith & D.M. Jones (Eds.), *Handbook of Human Performance*, Vol. 3. London: Academic Press.

Davis, D.R. (1948). *Pilot Error*. London: HMSO.

Folkard, S., & Åkerstedt, T. (1989). Towards the Prediction of Alertness on Abnormal Sleep/Wake Schedules. In A. Coblenz (Ed.), *Vigilance and Performance in Automated Systems*. Dordrecht: Kluwer.

Frankenhaeuser, M., & Johansson, G. (1981). On the Psychophysiologic Consequences of Understimulation and Overstimulation. In L. Levi (Ed.), *Society, Stress and Disease, Vol. 4 Working Life*. New York: Oxford University Press.

Fröberg, J., Karlsson, C.G., Levi, L., & Lidberg, L. (1975). Circadian Variations of Catecholamine Excretions, Shooting Range Performance and Self Ratings of Fatigue during Sleep Deprivation. *Biological Psychology*, *2*, 175-188.

Haslam, D.R. (1982). Sleep Loss, Recovery Sleep and Military Performance. *Ergonomics*, *25*, 163-178.

Hockey, G.R.J. (1986). Changes in Operator Efficiency as a Function of Environmental Stress, Fatigue and Circadian Rhythms. In K. Boff, L. Kaufman & J.P. Thomas (Eds.), *Handbook of Perception and Performance*, Vol. 2, 44/1-44. New York: Wiley.

Hockey, G.R.J. (1997). Compensatory Control in the Regulation of Human Performance under Stress and High Workload: A Cognitive Energetical Framework. *Biological Psychology*, *45*, 73-93.

Hockey, G.R.J. & Wiethoff, M. (1989). Assessing Patterns of Adjustment to the Demands of Work. In S. Puglisi-Allegra & A. Oliverio (Eds.), *Psychobiology of Stress*. Dordrecht: Kluwer.

Holding, D.H. (1983). Fatigue. In G.R.J. Hockey (Ed.), *Stress and Fatigue in Human Performance*. Chichester: Wiley.

Jongman, L., Meijman, T.F., & De Jong, R. (1999). Strategy Choice when being Mentally Fatigued. In: D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics*, Vol. 4. Ashgate.

Melamed, S. & Bruhis, S. (1996). The Effects of Chronic Industrial Noise Exposure on Urinary Cortisol, Fatigue and Irritability: A Controlled Field Experiment. *Journal of Occupational and Environmental Medicine*, *38*, 252-256.

Meijman, T.F. (1997). Mental Fatigue and the Efficiency of Information Processing in Relation to Work Times. *International Journal of Industrial Ergonomics*, *20*, 31-38.

Meijman, T.F., Mulder, G., Dormolen, M. van, & Cremer, R. (1992). Workload of Driving Examiners: A Psychophysiological Field Study. In H. Kragt (Ed.), *Enhancing Industrial Performance*. London: Taylor & Francis.

Muscio, B. (1921). Is a Fatigue Test Possible? (A Report to the Industrial Fatigue Research Board). *British Journal of Psychology*, *12*, 31-46.

Pearson, R.G. (1957). Scale Analysis of a Fatigue Checklist. *Journal of Applied Psychology*, *41*, 186-191.

Rabbitt, P.M.A. (1981). Sequential Reactions. In D.H. Holding (Ed), *Human Skill*. Chichester: Wiley.

Thorndike, E. (1900). Mental Fatigue. *Psychology Review*, *7*, 466-482.

Webster, D.M., Richter, L., & Kruglanski, A.W. (1996). On Leaping to Conclusions when Feeling Tired: Mental Fatigue Effects on Impressional Primacy. *Journal of Experimental Social Psychology*, *32*, 181-195.

Welford, A.T. (1965). Fatigue and Monotony. In O.G. Edholm & A.L. Backrush (Eds.). *The Physiology of Human Survival*. London: Academic Press.

Williams, H.L. Lubin, A., & Goodnow, J.L. (1959). Impaired Performance with Acute Sleep Loss. *Psychological Monographs*, *73*, 1-26.

3.2.5 Sleep Loss

3.2.5.1 Definitions and Measurement

Sleep loss is defined as any reduction in the amount of daily sleep obtained within a 24-hour day. For most individuals, this is a reduction to less than 6 to 8 hours of sleep per day, but individual differences in the daily sleep requirement can be considerable, ranging from extremes of approximately 4 hours to 12+ hours of sleep per 24 hours. The effect of sleep loss is increased sleepiness, which is the intervening variable mediating resulting decrements in (1) cognitive performance capacity, and (2) the continued ability to sustain wakefulness. It is not clear whether the relationship between sleep loss and performance decrements is linear or non-linear, but it is clear that greater levels of sleep debt are associated with greater levels of impairment, and it is clear that performance continues to be mediated to a significant extent by the endogenous circadian rhythm of alertness, even following extended sleep deprivation.

Of the various sleep parameters that are obtained with standard polysomnographic measurement techniques (EEG, EOG, EMG) and scoring (Rechtschaffen & Kales, 1968), only one parameter – *sleep duration* – has been shown to account for a significant portion of the variance in the recuperative value of sleep (Wesensten, Balkin, & Belenky, 1999). Precise measurement and quantification of sleep/sleep loss in the operational environment requires long-term (several days or weeks) monitoring to determine the typical sleep duration of the individual operator (to establish an individual baseline). Less precise but useful estimations of sleep loss can also be determined by comparing shorter-term measurements of sleep duration (e.g., only during actual operations) to population norms.

3.2.5.2 Background

The capacity to perform a specific cognitive task ultimately depends on the underlying capacity and readiness of the brain to perform that task. Normal performance over extended periods of time typically reflects and signifies a normal underlying level of brain functioning (e.g., normal alertness levels and an absence of pathologies or other stresses such as sleep loss). Also, normal performance typically involves some variability, with circadian (as well as ultradian) rhythmicity evident for performance of those tasks sensitive to fluctuations in alertness/sleepiness.

In those situations (such as sleep loss) in which brain functioning is compromised, the average performance level will typically be reduced to an extent that corresponds to the extent of the underlying brain dysfunction. This correspondence may not be perfect or linear since compensatory mechanisms such as increased focusing of concentration and effort may help maintain performance at nominally adequate levels, at least temporarily (Hockey, 1998). However, extended monitoring of performance (or more extensive probing of performance capacity with challenging tasks) will typically reveal deficits that reflect the compromised brain state.

Normal night-time sleep (e.g., from 11 p.m. to 6 a.m.) is necessary for maintaining alertness during the following day. Any deviation from the normal daytime wakefulness/night-time sleep pattern may cause fatigue and increase the risk of accidents. Although it is not possible to precisely determine the extent to which sleep loss and circadian rhythm factors may have contributed, it is notable that several catastrophic accidents over the past two decades have occurred during the descending phase – or near the nadir – of the circadian rhythm of alertness, including the partial meltdown of the nuclear power plant at Three Mile Island in 1979 (4 a.m.), the 1989 oil spill by the super tanker, Exxon Valdez (just after midnight), the core meltdown at Chernobyl in 1986 (1:23 a.m.); and the gas leak from a pesticide plant in Bhopal in 1984 (just after midnight).

3.2.5.3 Effects on Performance

Sleepiness constitutes a state of compromised brain functioning that is ubiquitous in the modern military operational environment, especially since these operations are now typically conducted on a 24-hour-per-day basis, and often entail surges during the night-time (to take advantage of what are typically superior night-vision and other imaging technologies).

It has long been known that sleep deprivation has a generally negative effect on psychomotor performance (the first scientific study of sleep deprivation on human performance was conducted in 1896 by Patrick and Gilbert). However, not all tasks are equally sensitive to sleep loss. In general, tasks involving mental performance are sensitive, whereas tasks requiring mostly physical performance (e.g., tasks mostly dependent upon muscular strength or endurance) are virtually impervious to sleep loss. Additionally, tasks involving higher-order mental abilities (i.e., those mediated by the prefrontal cortices such as reasoning, judgment, creativity, situational awareness, divergent thinking, and the ability to devise and execute appropriate multi-step plans of action) may be especially sensitive to sleep loss (e.g., Horne, 1988).

Also, performance has been shown to be reduced on tasks that are themselves sleep-conducive, including tasks that are of long duration, uninteresting, complex, or for which no feedback is provided (Wilkinson, 1965). However, this does not mean that performance decrements on such tasks are invariably the result of frank sleep onset. Although performance deficits following sleep loss can *sometimes* be attributed to "lapses", those momentary deficits in attention or other mental abilities (Lubin, 1967) that are occasionally associated with "microsleeps" (i.e., 1-10 sec episodes during which sleep stage 1-like EEG may be evident; Dement, 1972), sleepiness can result in performance decrements even during objectively (polysomnographically) verified wakefulness (e.g., Valley & Broughton, 1983; Balkin et al., 2000).

Both speed and accuracy of performance can be negatively affected by sleep loss, but the speed with which cognitive work is completed generally declines to a greater extent (Williams & Lubin, 1967). In fact, performance accuracy is often preserved at the expense of speed on self-paced tasks (i.e., a speed/accuracy trade-off often results from sleep loss; Lubin, 1967).

3.2.5.4 Assessment Methods

The sleep parameter that accounts for virtually all of the variance in recuperative value (and thus for virtually all of the variance in post-sleep performance capacity and alertness) is sleep duration (e.g., Wesensten et al., 1999). Thus, any accurate measure or correlate of sleep duration provides the information necessary to predict the operator's alertness state and associated cognitive performance capacity. However, it is important to note that the accuracy of sleep/wake history-based predictions of OFS depends not only on the accuracy of the sleep duration measurement tool itself, but also on the number of consecutive days or weeks over which measurements are continuously obtained. Longer data collection periods produce better predictions since they facilitate the assessment of individual differences in sleep needs and/or the detection of chronic levels of sleep debt.

3.2.5.4.1 Subjective Measures of Sleep and Alertness

Questionnaires, Rating Scales, and Sleep Diaries

Subjective estimates of sleep duration and resulting somnolence/alertness can be obtained using sleep questionnaires/diaries and sleepiness scales, respectively, although the validity and reliability of these measures when used for extended periods is unknown. There are a substantial number of sleep questionnaires and logs currently available, but the Stanford Sleepiness Scale (SSS) has been the standard measure of subjective sleepiness for many years (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973). The individual being tested selects one of seven statements describing different levels of extant sleepiness to describe his/her present state. Advantages of the SSS include its brevity and ease of administration and the fact that it can be completed repeatedly, which is also useful for evaluating circadian rhythm influences on sleepiness. On the other hand, the Epworth Sleepiness Scale (ESS; Johns, 1991), which describes the drive to sleep rather than sleepiness, and in which sleepiness is assessed more as a trait than a state (in order to determine the likelihood of sleep disorder), would be less useful in the operational environment since it is less sensitive to fluctuations in alertness over brief time intervals.

3.2.5.4.2 Physiological Measures of Sleep and Alertness

Polysomnography

Polysomnography, which requires, at minimum, EEG measurement from C3 or C4 sites, EOG, and facial EMG, is the "gold standard" for the characterization of human sleep. The defining criteria for identifying sleep onset and offset, and the various stages of sleep (1, 2, 3, 4, and REM) are delineated in the scoring manual produced by Rechtschaffen and Kales (1968). Currently, polysomnographic variables are most often obtained and recorded digitally, but scoring of sleep is performed by a human scorer who assesses the raw data displayed on a computer monitor in (typically) 30-second epochs. Automated sleep-scoring programs are available, but none has been adequately validated and none are currently sanctioned.

Although polysomnography can be obtained with currently available ambulatory recording systems, and high impedance/low maintenance electrodes can be used with these systems (which would reduce the attention and care that is needed to ensure adequate, artifact-resistant recordings), long-term assessment of sleep in large numbers of operators in the operational environment with polysomnography will remain prohibitively expensive in terms of direct costs (for the recording equipment and supplies) and staff requirements (for maintaining the equipment and assessing the data) for the foreseeable future.

3.2.5.5 References

Balkin, T., Thorne, D., Sing, H., Thomas, M., Redmond, D., Wesensten, N., Williams, J., Hall, S., & Belenky, G. (2000). *Effects of Sleep Schedules on Commercial Driver Performance* (Report No. DOT-MC-00-133). Washington, DC: U.S. Department of Transportation, Federal Motor Carrier Safety Administration.

Dement, W.C. (1972). Sleep Deprivation and the Organization of Behavioral States. In C.D. Clemente, D.P. Purpura, & F.E. Mayer (Eds.), *Sleep and the Maturing Nervous System*. New York: Academic Press.

Hockey, G.R.J. (1998). Effects of Sleep Deprivation and User Interface on Complex Performance: A Multilevel Analysis of Compensatory Control. *Human Factors*, *40*, 233-253.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W.C. (1973). Quantification of Sleepiness: A New Approach. *Psychophysiology*, *10*, 431-436.

Horne, J.A. (1988). *Why We Sleep: The Functions of Sleep in Humans and Other Mammals*. Oxford: Oxford University Press.

Johns, M.W. (1991). A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep*, *14*, 540-545.

Lubin, A. (1967). Performance under Sleep Loss and Fatigue. In S.S. Kety, E.V. Evarts, & H.L. Williams (Eds.), *Sleep and Altered States of Consciousness* (pp. 506-513). Baltimore: Williams and Wilkins.

Patrick, G.T.W., & Gilbert, J.A. (1896). On the Effects of Loss of Sleep. *Psychological Review*, *3*, 469-483.

Rechtschaffen, A., & Kales, A. (Eds.). (1968). *A Manual of Standardized Terminology, Techniques, and Scoring System for Human Subjects* (NIH Publication No. 204). Washington, DC: U.S. Government Printing Office.

Valley, V., & Broughton, R. (1983). The Physiological (EEG) Nature of Drowsiness and its Relation to Performance Deficits in Narcoleptics. *Electroencephalography and Clinical Neurophysiology*, *55*, 243-251.

Wesensten, N.J., Balkin, T.J., & Belenky, G. (1999). Does Sleep Fragmentation Impact Recuperation? A Review and Re-Analysis. *Journal of Sleep Research*, *8*, 237-245.

Wilkinson, R.T. (1965). Sleep Deprivation. In O.G. Eldholm & A.L. Bacharach (Eds.). *The Physiology of Human Survival* (pp. 399-430). London: Academic Press.

Williams, H.L., & Lubin, A. (1967). Speeded Addition and Sleep Loss. *Journal of Experimental Psychology*, *73*, 313-317.

3.3 TASK CHARACTERISTICS

3.3.1 Cognitive Load

3.3.1.1 Definitions and Measurement

Cognitive load reflects the mental activity that is involved in the processing of task-relevant information. It depends upon the amount and quality of information to be processed and the current capacity of the operator. Cognitive load is an important aspect of the concept of mental workload, which can be defined as "*an intervening variable similar to attention that modulates or indexes the tuning between the demands of the environment and the capacity of the operator*" Kantowitz (1988). This definition highlights the two main features of mental workload within the human factors research of the last few decades (i.e., the capacity of the operator and the task demands). Mental workload is high when the task demands are high and/or the capacity of the operator is low. The capacity of an operator is not fixed, but can be moderated by environmental and individual factors (see other sections). There are several aspects of a task that can make it more cognitively demanding. Neerincx, van Doorne, and Ruijsendaal (2000) provide a model in which three important task factors are incorporated:

1. *Time pressure*. This is the time that is required to perform a task in relation to the available time. Workload will become higher when the difference between the required and the available time is small.

2. *Task set switches*. When a new task is performed, the operator must retrieve information from different sources and build a mental model for the task. Every time an operator switches from one task to another he/she must switch between different mental models. This requires the operator to keep multiple mental models active. This process leads to additional information processing, which increases the workload. Rapid switching between tasks further increases the workload.

3. *Level of information processing*. Tasks can be performed at several cognitive levels. Rasmussen (1986) distinguishes three levels of information processing: target-oriented *skill-based* behavior at the lowest level, procedure-oriented *rule-based* behavior at the intermediate level, and goal-controlled *knowledge-based* behavior at the highest level. Skill-based tasks are well-trained tasks that require minimal attention, such as automobile driving by an experienced driver. Rule-based tasks are performed in accordance with specific rules. Standard procedures can be executed after information is retrieved. For example, when you see the traffic light turn red, stop your car. The influence upon workload depends on the level of training. Highly trained tasks require only modest attention and therefore have little effect on workload. Knowledge-based tasks involve mainly planning and management. They require (1) high-level situation assessment of unfamiliar situations for which no rules are available from previous encounters, and (2) the consideration of alternative actions at a strategic level. These tasks require much attention and therefore strongly affect mental workload. Workload will often be affected by the quantity of knowledge-based tasks.

Norman and Bobrow (1975) introduced another relevant aspect of the relation between task demands and workload. Tasks can be "resource limited" or "data limited". For resource-limited tasks, performance depends upon mental resources or mental effort expenditure. Performance will improve when the operator invests more effort. An example is a tracking task in which one can increase performance by expending more effort. Performance on data-limited tasks depends upon the quality of the information and not upon the effort investment. For example, if you are to keep an aircraft at a fixed altitude using a poor altimeter, spending additional effort will not improve performance.

Both resource-limited and data-limited tasks will increase workload subjectively. In both situations it is difficult to get an adequate level of performance, and operators will interpret this as a high workload situation. Only resource-limited tasks will affect objective (physiological) workload measures.

3.3.1.2 Background

During the past century, the nature of most work has changed from mainly physical to primarily mental work. Information processing has become the most important aspect of task performance for many jobs.

3.3.1.3 Effects on Performance

The effect of workload upon task performance is illustrated in Figure 4. When the task demands are low (low workload), the relationship between demands and performance is ambiguous. For short-lasting tasks, operators can easily perform at a maximal level. However, on some occasions (e.g., vigilance situations), performance can become very low. Vigilance is required when operators are expected to pay continuous attention to a task (high time pressure), the task lasts longer than about 20 minutes, and the required level of information processing is low (rule-based or skill-based).

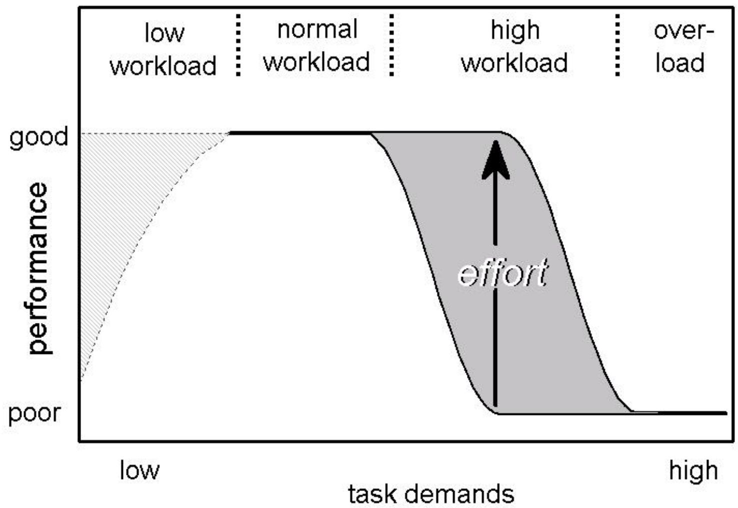

Figure 4: Relation between Workload, Task Demands, Performance, and Effort (See text for explanation).

Under normal workload conditions, performance will typically be high. Under high workload conditions, performance will depend upon the level of effort expenditure. When operators do not invest the necessary effort, performance will decline rapidly as task demands increase. When operators increase the effort investment, the level of performance can remain at a maximal level during a broad range of task demands. However, operators may not always strive for maximal performance. They will constantly regulate their effort expenditure in an efficient manner. Additional effort will only be invested for a substantial improvement in performance. When the relationship between the amount of effort expenditure and performance improvement is weak, such as for data-limited tasks, operators may not invest the additional effort.

In a situation of overload, performance will be poor regardless of the level of effort expenditure.

3.3.1.4 Assessment Methods

There is no clear definition of cognitive load. Thus, there does not exist a general assessment technique that is able to capture the entire concept of cognitive load. The present assessment methods can be divided into three categories: (1) performance measures, (2) subjective rating scales, and (3) physiological measures. These measures assess different aspects of cognitive load. Furthermore, many assessment techniques are affected by factors other than cognitive factors. Therefore, it is very important to have information about what each measure can provide and what it does not provide. The information below states what each measurement category can tell us about cognitive load.

3.3.1.4.1 Performance Measures

Information about performance is very important because one of the main reasons for assessing operator state is to obtain information about possible breakdowns in performance. Performance measures can be divided into primary and secondary measures. Primary task measures are directly related to the main tasks that are being performed. Secondary task measures are intended to measure the spare capacity of the operator by presenting an additional task. The concept of secondary tasks is that the mental workload is acceptable when operators have available spare capacity and are able to perform a secondary task. Another approach is to measure the performance of embedded secondary tasks. Embedded secondary tasks are part of the overall task load but have an established lower priority. For example, pilots may ignore radio communications when the primary flying task difficulty is high. By measuring their responses to radio calls, it is possible to determine if the primary task workload is high.

Figure 4 shows that performance provides relevant information only when examined in combination with the level of mental effort expenditure. A low level of performance can be due to a high cognitive load or to low effort expenditure. For example, when an operator has enough spare capacity but is not motivated to expend additional effort, performance will be low under high workload conditions. On the other hand, when an operator is eager to expend effort, performance can be high even under high workload conditions.

3.3.1.4.2 Rating Scales

Rating scales are the most commonly used cognitive load assessment tool, mainly because they are easy to use, relatively cheap, and do not require special equipment. They have high operator acceptance because they provide the operator the opportunity to give opinions about the system. They have been found to be very sensitive to changes in task load. However, they are not always diagnostic. For example, Veltman, Gaillard, and van Breda (1997) showed that pilots tend to give high mental effort ratings when their performance decreases, even when they have not really invested extra effort. Ratings are also subject to operator bias and may be affected by memory if not obtained immediately after task performance. For more information, see the section on subjective assessment methods in this report.

3.3.1.4.3 Psychophysiology

Psychophysiological measures provide objective information about cognitive load (Kramer, 1991; Wilson & Eggemeier, 1991). Because they primarily reflect effort expenditure, they are only sensitive in the region of high workload. Aasman, Mulder, and Mulder (1987) showed that heart rate variability decreases when subjects spend more effort, but when the task demands exceed operator capacity and operators do not remain engaged in the task, heart rate variability increases. Heart rate has had the most widespread use as a measure of mental workload. Generally, heart rate increases and the variability of the heart rhythm may decrease with increased task demands. Changes in heart activity are generally not diagnostic about the cause of the increased workload but serve as an index that the level has changed (Wilson, 1993). Eye blink rate has been reported to be diagnostic of increased visual demand. Blink rate typically decreases with increased visual demand (Veltman & Gaillard, 1998). EEG changes have also been noted. Most often the power in the posterior alpha band decreases. Also, increased midline frontal theta has been reported with increased task demands (Sterman, Mann, Kaiser, & Suyenobu, 1994; Wilson, 2002). See the relevant sections of this report for further information on each of the psychophysiological measures mentioned here.

It is necessary to take into account that the accuracy of such measures depends upon the individual's psychophysiological "norm" (Karpenko et al., 1984). The psychophysiological measures also require special equipment and analysis procedures as well as trained personnel. They are susceptible to various artifacts that must be detected and removed.

3.3.1.5 References

Aasman, J., Mulder, G., & Mulder, L.J.M. (1987). Operator Effort and the Measurement of Heart Rate Variability. *Human Factors*, *29*, 161-170.

Kantowitz, B.H. (1988). Mental Workload. In P.A. Hancock (Ed.), *Human Factors Psychology* (pp. 81-119). Amsterdam: Elsevier.

Karpenko, A., Burov, A., Kalnish, V., Kapshuk, A., Grigorus, A., Grigorjantz, T., & Bobko, N. (1984). Control of Current Reliability of Operators' Work on the Basis of Psychophysiological Criteria (in Russian). *Nuclear Power Plant*, *7*, 156-162.

Kramer, A.F. (1991). Physiological Metrics of Mental Workload: A Review of Recent Progress. In D. Damos (Ed.), *Multiple-Task Performance* (pp. 279-328). London: Taylor and Francis.

Neerincx, M.A., van Doorne, H., & Ruijsendaal, M. (2000). Attuning Computer-Supported Work to Human Knowledge and Processing Capacities in Ship Control Centres. In J.M.C. Schraagen, S.E. Chipman, & V.L. Shalin (Eds.), *Cognitive Task Analysis* (pp. 341-362). Mahwah, NJ: Erlbaum.

Norman, D.A., & Bobrow, D.G. (1975). On Data-Limited and Resource-Limited Processes. *Cognitive Psychology*, *7*, 44-64.

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. Amsterdam: Elsevier.

Sterman, M.B., Mann, C.A., Kaiser, D.A., & Suyenobu, B.Y. (1994). Multiband Topographic EEG Analysis of a Simulated Visuomotor Aviation Task. *International Journal of Psychophysiology*, *16*, 49-56.

Veltman, J.A., & Gaillard, A.W.K. (1998). Physiological Workload Reactions to Increasing Levels of Task Difficulty. *Ergonomics*, *41*, 656-669.

Veltman, J.A., Gaillard, A.W.K., & van Breda, L. (1997). Workload Indices: Physiological Measures versus Subjective Ratings. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics. Volume One: Transportation Systems* (pp. 269-275). Aldershot: Ashgate.

Wilson, G.F. (1993). Air-to-Ground Training Missions: A Psychophysiological Workload Analysis. *Ergonomics*, *36*, 1071-1087.

Wilson, G.F. (2002). An Analysis of Mental Workload in Pilots during Flight using Multiple Psychophysiological Measures. *International Journal of Aviation Psychology*, *12*, 3-18.

Wilson, G.F., & Eggemeier, F.T. (1991). Physiological Measures of Workload in Multi-Task Environments (pp. 329-360). In D. Damos (Ed.), *Multiple-Task Performance*. London: Taylor and Francis.

3.3.2 Physical Load

3.3.2.1 Definition

Exercise is a process whereby the body performs work. Physical activity is classified according to the physiological characteristics of the movements that are performed. It is customary to classify the work performed by human muscles as dynamic (isotonic) and static (isometric; Shepard, 1972; Astrand & Rodahl, 1970).

Dynamic work involves the shortening and lengthening of specific muscles and is associated with action to change the body position in space or the positions of parts of the body with respect to each other. The work is accomplished by repetition of a cycle of movements with alternating phases of muscle contraction and relaxation.

Static work is performed by a muscle when the position of the body or its parts is kept constant. During static work, the muscle expends energy to maintain tension, which serves to maintain body posture in a gravitational field. A strongly tensed muscle exerts considerable internal pressure that compresses the blood vessels and reduces the blood flow within active muscles, which may elicit pain and cause fatigue. Under typical conditions, human motor activity consists of a complex combination of dynamic and static work.

The work rate (the amount of work performed per unit time) depends on the force, amplitude, and frequency of muscular contraction. The maximum time for which a given type of work can be performed depends on the relationship of work rate to the maximum power output of the person. If a high proportion of power is used, the rapid onset of fatigue will allow only a short duration of work. However, if the work rate is low, activity can continue for a long time without undue fatigue.

Cyclic physical activity can be classified according to the duration and speed of performance. Brief efforts may be maximal; however, light activity over a long period may also be fatiguing.

3.3.2.2 Background

The performance of physical work depends upon the conversion of the chemical energy in adenosine triphosphate (ATP) and creatine phosphate (CP) into mechanical work using the transducing system of the skeletal muscles including the long-chain proteins actin and myosin (Falls, 1968; Edington & Edgerton, 1976). The maximum force that can be generated in static effort depends on muscle characteristics, while the duration for which the contraction can be sustained is determined by personal reactions to the accumulation of acid metabolites generated in the synthesis of ATP and CP. The maximum power of steady dynamic effort depends on the ability to resynthesize ATP and CP using glycogen and fat stored within active muscle fibers. With brief periods of intensive work, the oxygen delivery system may be overtaxed and energy is then derived from anaerobic metabolism of stored glycogen. If the effort is continued for more than a few minutes, the main determinant of performance becomes the delivery of oxygen to the muscle fibers. If exercise is further continued for an hour or more, effort may in turn be limited by an accumulation of body heat, a depletion of fluid reserves through sweating, or an exhaustion of local glycogen and fat (Shepard, 1977). The metabolic rate of muscle is low at rest (approx. 3 ml.min-1 of oxygen per kg of tissue); however, during maximum aerobic activity, the muscles consume oxygen 100 times as rapidly as this, and a further brief 3-fold to 4-fold increase of power is possible by calling upon anaerobic energy-generating systems. However, the completeness of oxygen extraction evident in blood leaving the active muscle indicates that steady effort is being limited by oxygen flow within the cardiorespiratory delivery system.

At quiet wakefulness, the respiratory minute volume is about 4-7 $l/min/m^2$ of body surface area. Ventilation increases during exercise in parallel with oxygen consumption, although at more than 70% of maximum oxygen intake, the accumulation of anaerobic metabolites initiates a disproportional hyperventilation sometimes called the "anaerobic threshold".

In certain extreme types of physical performance, motivation and problems of fluid and heat balance can limit the rate of working. Except for such extreme cases, the most important determinants of heavy physical performance are: (1) for the first 10 seconds of effort, the maximum rate of anaerobic energy (anaerobic power), (2) for the period of 10 to 60 seconds, the tolerance to anaerobic metabolites (anaerobic capacity), and (3) for activities lasting longer than 1 minute, the delivery of oxygen to sustain the aerobic release of energy (aerobic power of maximum oxygen intake).

Energy expenditure at rest is approximately 2.8 kJ.min-1.m2 of body surface area in a young adult man, although it diminishes with age. It is lower per unit of body surface area in women because the body contains more fat. Anaerobic power reaches its maximum in jumping and throwing movements. Anaerobic capacity is limited by the accumulation of lactic acid in the blood and muscles.

Dynamic muscular exercise performed at a steady submaximal rate evokes oxygen consumption increases during the first few minutes of exercise, reaching a plateau (steady-state level) with a half time of 30 to 40 seconds. Adaptation proceeds faster in fit young subjects and more slowly in the elderly and persons with cardiorespiratory disease. Independent of the intensity of exercise, an oxygen deficit appears at the beginning of physical exercise due to a delay in circulatory adjustment to associated anaerobic conditions in the muscles at the beginning of activity. During recovery, oxygen consumption may remain above previous resting values for as long as one hour.

3.3.2.3 Effects on Performance

The capacity to perform a cognitive task depends on the individual's baseline physical fitness as well as on the intensity and duration of the physical load. Physical load leading to fatigue at operational work is often regarded as important because it may interfere with the high efficiency demanded in many occupational settings. Examples of possible detrimental consequences are poor judgment, omission of details, indifference to essentials, and generally inadequate performance (Schwab, 1953). In healthy subjects, fatigue is a normal phenomenon experienced by everyone and usually easily relieved by rest or sleep. However, fatigue that becomes excessive or chronic without recovery may lead to tiredness and exhaustion affecting OFS and performance. Fatigue has a complex nature and might be categorized into three types: (1) "physiological fatigue" as a reduction of physical capacity, (2) "objective fatigue" as a work decrement, and (3) "subjective fatigue" as feelings of weariness. These three types of fatigue may actually be defined in the following terms: as the physical capacity the person possesses, as the work the person achieves, and as the feelings the person possesses. This feature triad of fatigue has been widely recognized (Bills, 1934). Perceived fatigue is related to the work task being performed, and specific work tasks differ in the kind of demands they impose on a person.

In addition to physical workload, there are other conditions that also affect the general state of the person. These include mental load, sensory load, time of day, the psychological and physical environment, and individual characteristics. Physical load can be described as whole body work or local physical work (Kilbom, 1987). *Whole body work* consists of dynamic load on large muscle groups and makes demands on a person's oxygen uptake capacity. *Local physical work* often consists of low, steady loading of small muscle groups and makes demands on a person's capacity to develop and maintain muscle force. Both for static muscular loading and dynamic muscular work, endurance is related to the developed force as a proportion of the muscle's maximal force capability. The stronger the muscles, the greater the load they can endure without developing muscular fatigue (Astrand & Rodahl, 1986). Physical fatigue is often regarded as synonymous with muscular fatigue, but it has also been recognized as a more complex phenomenon influenced by both physiological and psychological, factors. The manifestation of fatigue may be described as reduction of physical capacity, as work reduction, and as feelings of weariness leading to a decrease in performance. Physical load leading to physical fatigue associated with moderate to heavy aerobic and anaerobic activity would result in an increased propensity for general or mental fatigue, thereby eliciting deterioration of operational performance.

3.3.2.4 Assessment Methods

Physiological measures play an important role in the assessment of physical load effects on OFS and operational performance. Different assessment methods and physiological variables can be used as indicators of physical load leading to fatigue. Local muscle fatigue can be quantified through disturbances at the cellular level by measuring biochemical and ionic changes (Vollestad & Sejersted, 1988), lactates (Gamberale, 1972), changes in electromyography (Hagberg, 1981; Malmquist, Ekholm, Lindstrom, Petersén, & Örtengren, 1981), and changes in blood pressure and heart rate (Bystrom, Mathiassen, & Fransson-Hall, 1991; Kilbom, Gamberale, Persson, & Anwall, 1983). As indicators of energy consumption, changes in heart rate and oxygen consumption have often been interpreted as general physical fatigue (Gamberale, 1972).

Different principles and types of exercise have been employed in physical workload tests. The objectives of the testing are (1) to test the operator's fitness for work and other activities; and (2) to assess the functional status of cardiovascular and respiratory functions. Exercise tests may be maximum tests in which exercise of increasing intensity is performed until no further increase in oxygen uptake occurs or submaximum tests in which exercise is performed at lower intensities of effort than maximum tests.

Work capacity may be assessed using the following indices from exercise tests:

- *Maximal power output* – the highest rate of work achieved during the test.
- *Endurance time* – total exercise time to exhaustion or to predetermined endpoints in a continuously graded test.
- *Physical working capacity* – the highest rate of work at which heart rate and respiratory rate do not exceed 170 beats/min and 30 breaths/min, respectively, during continuous graded bicycle exercise.
- *Total work* – accumulated work to exhaustion or to predetermined endpoints during cycle or treadmill tests.

Endurance tests may be used to study cardiorespiratory and/or metabolic effects at various intensities of effort. They can also be used to collect data in thermoregulatory and altitude studies and in other studies examining the effects of endurance exercise on blood concentrations of hormones, electrolytes, etc. The exercise intensity used in a protocol is usually determined as a percentage of one's VO_2 max; if so, VO_2 max must be determined before the endurance test can be performed. These two procedures, under almost all circumstances, are not conducted on the same day.

To determine VO_2 max, subjects undergo a maximal graded exercise test to voluntary exhaustion. This means that the subject makes the decision when the test is over. On a cycle ergometer, the test is terminated when the subject can no longer turn the cranks at the desired frequency; on the treadmill, the test is terminated when the subject can no longer run at the treadmill speed and stands straddling the treadmill belt while holding the railing. In addition, the subject is spotted at his/her side to prevent a possible fall. When testing an athlete in particular, the person is coached to proceed as long as possible.

Untrained individuals and older subjects are encouraged to "give a hard effort," but not coached to continue to their absolute physical limits. Following a test, the subject goes through a cool-down at a self-selected intensity until recovered.

Test protocols vary somewhat in terms of the duration of each stage (1-3 minutes) and increments in intensity, but all begin with a 5-10 minute warm-up followed by a gradual increase in work effort until volitional exhaustion. Typically, a graded exercise test takes 8-12 minutes, excluding warm-up and cool-down. Throughout the test, the subject breathes through a rubber mouthpiece and a two-way re-breathing valve that is connected by low-resistance tubing to a metabolic measurement system. A computer provides analysis of the expired air to determine oxygen consumption and carbon dioxide production. Heart rates are monitored continuously, usually with a heart rate monitor. Older, non-active subjects are also monitored for heart arrhythmias and/or ischemic changes using ECG. Blood pressure is also monitored and recorded at each stage with older subjects.

3.3.2.4.1 Physiological Measures

Oxygen Consumption (VO_2)

Maximal oxygen consumption (VO_2 max) is the best index of exercise capacity and a measure of the functional limit of the cardiovascular system to physical load. The testing of maximal aerobic power through direct measurement of VO_2 max is considered the best measure of cardiovascular fitness. VO_2 max is often used in studies to determine the effects of exercise training on fitness, both from short-term training (e.g., several weeks) to longitudinal studies of a year or longer. This testing procedure involves exercise of increasing intensity until oxygen consumption reaches a plateau, which is the criterion indicating that the maximum level has been reached. The direct determination of VO_2 max intake requires that the subject performs vigorous exercise at a maximum or hypermaximum level and that oxygen intake is actually measured. Lactates may be measured as a marker of shifting to an anaerobic level of exercise.

Heart Rate (HR)

Indirect assessment of energy expenditure, or simply of physical activity, may be attempted using heart rate (HR) because there is a general relationship between HR and oxygen consumption. This method is most precise at high levels of energy expenditure reaching 50-90% of maximum oxygen intake. This relationship provides a basis for the monitoring of physical activity by recording the HR.

HR has also been one of the most popular psychophysiological measures to monitor OFS. It is easy to measure and has been found to be very sensitive to changes in operator state.

An increase in HR due to a decrease in parasympathetic control is an immediate response of the cardiovascular system to exercise. This increase in HR is followed by an increase in sympathetic control to the heart and systematic blood vessels. During exercise, HR increases linearly with workload and VO_2. During mild work at a constant work rate, HR reaches steady state within several minutes. As workload increases, the time necessary for HR to stabilize will progressively lengthen. Recovery of HR after exercise is dependent on the baseline level of fitness of the subject.

Relatively rapid HR during submaximal exercise or recovery could be due to deconditioning. An inadequate rise or fall in systolic BP during exercise can occur. Some normal subjects have a transient drop in systolic BP at maximum exercise or immediately after exercise. The HR response to exercise and the recovery time after exercise depend on the baseline level of autonomic HR control, which is strongly related to the subject's fitness.

Respiration

The respiratory minute volume is approximately 4 l/min per m^2 of body surface area (about 7 l/min in a young man). During physical exercise, ventilation increases, at first in parallel with oxygen consumption. At more than 70% of maximum oxygen intake, the accumulation of anaerobic metabolites initiates a disproportionate hyperventilation. This is called the “anaerobic threshold”. In maximum exercise, a young untrained man may develop a respiratory minute volume of approximately 100 l/min. The respiratory rate increases from about 14 breaths/min to 40-50 breaths/min, while the tidal volume increases from 10% to 50% of the vital capacity, mainly at the expense of the inspiratory reserve.

The increased respiration during exercise secures a normal or a slightly increased alveolar oxygen tension, which in moderate work is also unchanged. However, it can show a small reduction with the widened alveolar-arterial gradient of vigorous exercise, but falls with intensities of effort that demand anaerobic metabolism. If vigorous physical work is undertaken, anaerobiosis supplements the aerobic processes in providing energy for muscular contraction and lactic acid is formed. The threshold for the initiation of anaerobic metabolism depends on the physical fitness of the person and on the type of exercise that is performed (typically 50-60% of maximum oxygen intake on a bicycle ergometer and 70-80% on a treadmill because of broader distribution of the task across the body musculature). With moderate effort, muscle vasodilation and a progressive increase of systemic pressure may correct the early build-up of lactate, but with heavier exercise, lactate continues to accumulate until the person is forced to stop exercising because of weakness and pain in the active muscles. The accumulation of lactic acid causes metabolic acidosis, provoking a disproportionate hyperventilation.

Identification of the lactate threshold is the best predictor of performance over a range of endurance distances. Training causes a shift in the exercise intensity at which the lactate threshold occurs, thus this test can be used to monitor the training/detraining progression.

Blood Pressure (BP)

BP is dependent on cardiac output and total peripheral resistance. Systolic BP rises with increasing dynamic workload as a result of increasing cardiac output. Diastolic BP usually remains about the same or may decline toward zero in some normal subjects, especially in well-trained sportsmen. Changes in BP reflect more than the contractile function of the left ventricle since they also depend on peripheral resistance. A drop in systolic BP below standing rest is of great concern during exercise or recovery. When exercise is terminated abruptly, some healthy persons have precipitous drops in systolic BP due to venous pooling. After maximum exercise there is usually a decline in systolic BP, which normally returns to the resting level in 6 minutes, then often remains lower than pre-exercise levels for several hours.

Systolic BP at maximum exertion or at immediate cessation of exertion is considered a useful first approximation of the heart’s isotropic capacity.

Blood Flow

Resting muscle blood flow is low (2-4 $ml.min^{-1}$ per 100 g of tissue). During rhythmic exercise, the diffusion pathway for oxygen within the tissue is shortened by at least a 3-fold increase in the number of capillaries. Local blood flow to active muscles increases roughly in direct proportion to the work being performed. When the physical activity is sustained by a smaller group of muscles, the direct restriction of perfusion by the contracting muscles may lead to some decrease of local blood flow at the highest rates of work. This phase is associated with a rise in systemic pressure (which might be expressed much more than in static work) and an accumulation of lactic acid from reliance upon anaerobic metabolism. Blood flow is first impeded when a muscle contracts at more than 15% of its maximum contractile force and the vessels are completely blocked at greater than 70% of maximum force. For example, hard work on a bicycle ergometer develops 25-35% of maximum force.

Pulse Oximetry

Pulse oximetry provides estimates of arterial oxyhemoglobin saturation (SaO_2) by utilizing selected wavelengths of light to non-invasively determine the saturation of oxyhemoglobin (SpO_2). Pulse oximetry can be performed by trained personnel in a variety of settings including operational performance. Oxymetry is appropriate for continuous and prolonged monitoring (e.g., during sleep, exercise, or operator activity). Exercise testing may be performed to determine the degree of oxygen desaturation and/or hypoxemia that occurs on exertion.

Electromyography (EMG)

EMG is used to measure the electrical signal associated with the activation of muscle tissue, whether a voluntary or involuntary contraction is involved. The EMG activity of voluntary muscle contractions is related to the muscle tension. Various EMG applications can be described. Diagnostic EMG involves examination of the characteristics of the motor unit action potential for duration and amplitude. These studies are typically conducted to help diagnose neuromuscular pathology. They also evaluate the spontaneous discharges of relaxed muscles and are able to isolate single motor unit activity. Kinesiological EMG is used primarily for movement analysis. This EMG application examines the relationship of muscular function to movement of the body segments, and evaluates timing of muscle activity with regard to the movements. Additionally, EMG has been used in attempts to examine the strength and force production of the muscles themselves, which may be useful for evaluating muscle function under physical load.

Hormonal Responses

Vigorous exercise induces responses in the endocrine system. Among other functions, these responses help to adjust the supply of metabolic fuels, stabilize the circulation, conserve fluid, and (in a more long-term sense) favor muscle hypertrophy and the conservation of essential mineral ions.

Exercise leads to an increased output of adrenal corticosteroids. In some cases, this increase might be due not only to exercise but to the associated emotional stress. Cortisol inhibits hexokinase and this helps to stabilize blood glucose level. Disease of the adrenal glands leads to marked weakness and fatigue. Repeated exercise, particularly in hot climates, increases the output of aldosterone, exerting a sodium conserving action. Moderate exercise does not change the blood level of adrenaline and noradrenaline, although exhausting work leads to a 2-fold to 3-fold increase in plasma noradrenaline derived from sympathetic nerve terminals rather than the adrenal glands together with a substantial decrease of plasma adrenaline.

Sustained and vigorous exercise leads to an increase in the blood level of the anterior pituitary growth hormone, with a peak concentration after one hour of continued effort. The growth hormone may facilitate muscle hypertrophy. It also inhibits the phosphorylation of glucose by the enzyme hexokinase and increases the mobilization of fatty acids, thus conserving carbohydrates during sustained work.

Vigorous exercise inhibits urine formation through an enhanced secretion of the anti-diuretic hormone of the posterior pituitary gland.

Blood insulin decreases and plasma glucagons increase during exercise, these changes being reversed during recovery. Insulin facilitates the action of hexokinase, and thus the uptake of blood glucose by the active muscles. Although the blood concentration of insulin falls during exercise, this drop is more than compensated by the increase in muscle blood flow, so that active muscle shows a substantial increase in the arterio-venous glucose difference during acute bouts of physical activity. Regular exercise reduces the insulin need of the diabetic because muscle glucose uptake for a given blood insulin level increases during exercise.

There is some evidence that exercise increases blood testosterone levels, and this may facilitate muscle hypertrophy in athletes. There have been documented changes in the performance of women throughout the menstrual cycle. Skilled performance seems to deteriorate during the phase of premenstrual tension, and some tasks are performed slightly better than normal during the phase involving menstrual flow.

3.3.2.4.2 Subjective Measures

For subjective assessment of experienced physical load, some subscales of the NASA TLX (Task Load Index) can be used. Self-reported symptoms as signs of physical fatigue ("exertion", "discomfort", and "aching") are also quite valuable (Ljunggren, 1985).

The NASA TLX was developed to assess mental workload in space and aerospace applications. The Computer Assisted Subjective Workload Assessment version of the NASA TLX is a fully automated revision of its predecessor pencil and paper tool, which experimental participants completed after they had finished the task being evaluated. The tool calculates the total amount of workload experienced and provides an indication of its source. The NASA TLX is a multidimensional subjective rating procedure that provides an overall workload score based on a weighted average of ratings on six subscales: (1) Mental Demands, (2) Physical Demands, (3) Temporal Demands, (4) Own Performance, (5) Effort; and (6) Frustration. Three dimensions relate to the demands imposed on the subject (Mental, Physical, and Temporal Demands) and three dimensions relate to the interaction of the participant with the task (Effort, Frustration, and Performance). An overall weighted measure of task load is also calculated on the basis of the scales. The NASA TLX is assumed to have acceptable diagnosticity because it uses six subscales that can be analyzed separately. The TLX has been tested in a variety of experimental settings that range from simulated flight to supervisory control simulations and laboratory tasks. The results of the first validation study were summarized by Hart and Staveland (1988). The derived workload scores have been found to have substantially less between-rater variability than uni-dimensional workload ratings, and the subscales provide diagnostic information about the sources of load.

3.3.2.5 Appropriateness for Measuring

The measurement of maximum oxygen consumption is accepted as the "gold standard" and is finding increased acceptance in the evaluation of OFS as well as for sport, exercise prescription, clinical practice, and the management of training and rehabilitation programs. Since there is a considerable scatter of "normal values", the data are best used to indicate the current work tolerance of the subject and his subsequent progress without reference to the supposed normality of subject's status.

Aerobic capacity and maximum power remain difficult to measure accurately, and since they seem of less significance for everyday activities, they are not used in the normal examination of work tolerance.

Because oxygen consumption is linearly related to HR frequency, changes in the latter might be the most acceptable for subjects undergoing high levels of exercise in daily practice. HR measurement during the tests or performance is a very simple procedure not evoking any discomfort. HR responses to physical load depend on the baseline level of autonomic HR control.

The ease of use of physical load tests is the most attractive in the testing of baseline operator functional (physical) capacity.

3.3.2.6 References

Astrand, P.O., & Rodahl, K. (1970). *Textbook of Work Physiology*. New York: McGraw Hill.

Astrand, P.O., & Rodahl, K. (1986). *Textbook of Work Physiology. Physiological Bases of Exercise*. New York: McGraw Hill.

Bills, A. (1934). *General Experimental Psychology*. New York: Longmans, Green & Co.

Bystrom, S., Mathiassen, S., & Fransson-Hall, C. (1991). Physiological Effects of Micropauses in Isometric Handgrip Exercise. *European Journal of Applied Physiology*, *63*, 405-411.

Edington, D.W., & Edgerton, V.R. (1976). *The Biology of Physical Activity*. Boston: Houghton Mifflin.

Falls, H.B. (1968). *Physiology of Exercise*. New York: Academic Press.

Gamberale, F. (1972). Perceived Exertion, Heart Rate, Oxygen Uptake and Blood Lactate in Different Work Operations. *Ergonomics*, *15*, 545-554.

Hagberg, M. (1981). On Evaluation of Local Muscular Load and Fatigue by Electromyography. *Arbete och halsa*, *24*.

Hart, S.G., & Staveland, L.E. (1988). Development of a Multi-Dimensional Workload Rating Scale: Results of Empirical and Theoretical Research. In P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-184). Amsterdam: Elsevier.

Kilbom, A. (1987). Fysiskt arbete, fysiologisk belastning [Physical Work, Physiological Load]. In N. Lundgren, G. Luthman, & K. Elgstrand (Eds.), *Maninniskan i Arbete* (pp. 48-92). Stockholm: Nordstedts Forlag.

Kilbom, A., Gamberale, F., Persson, J., & Anwall, G. (1983). *European Journal of Applied Physiology*, *50*, 179-193.

Ljunggren, G. (1985). Studies of Perceived Exertion during Bicycle Ergometer Exercise – Some Applications. Unpublished Doctoral Thesis, Stockholm University.

Malmquist, R., Ekholm, I., Lindstrom, L., Petersén, I., & Örtengren, R. (1981). Measurement of Localized Muscle Fatigue in Building Work. *Ergonomics*, *24*, 695-709.

NASA (1986). NASA Task Load Index (TLX): Computerized Version. Moffet Field, CA: NASA Ames Research Center, Aerospace Human Factors Research Division.

Schwab, R. (1953). Motivation in Management of Fatigue. In W. Floyd & A. Welford (Eds.), *Symposium on Fatigue,* (81, pp. 671-700). London: H.K. Lewis & Co.

Shepard, R.J. (1972). *The Physiology of Physical Activity*. Springfield, IL: C.C. Thomas.

Shepard, R.J. (1977). *Endurance Fitness* (2nd ed.). Toronto: University of Toronto Press.

Volestad, N., & Sejersted, O. (1988). Biochemical Correlates of Fatigue. A Brief Review. *European Journal of Applied Physiology*, *57*, 336-347.

3.3.3 Situational Awareness

3.3.3.1 Background

At present there is no universally accepted definition of situation awareness (SA). However, most researchers quote Endsley's (1987, 1990) definition of situation awareness as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of the status in the near future." SA thus revolves around the human operator's awareness of current conditions and the incorporation of this information in their actions. This definition emphasizes the fact that the human operator needs SA, and that the possession and loss of SA are very real phenomena in the operational experience. Situational awareness appears to be more than just a convenient concept to explain why performance sometimes succeeds or fails. To human operators in action it is a very real component of their experience in the process of performing complex tasks in dynamic, high-risk situations. While the concept of SA is not consistently accepted in the literature, many people nevertheless know "what it is like" to have SA and "what it is like" to not have SA. In this vein, perhaps the simplest definition of SA is contained in a quote by a pilot "Knowing what's going on so you can figure out what to do" (Adam, 1993).

In other words, having good SA means being in a high state of readiness to adapt or revise your current course of action with respect to your given situation in order to succeed in meeting your ongoing operational objectives. Losing SA is akin to attempting to meet those objectives "blind", without knowing how your course of action may need to be adapted or changed. Realizing your SA is reduced or off-course provides the incentive to seek and acquire the currently available key information. Losing SA (having a false awareness of the situation) and not realizing it is even worse: it means you may proceed "confidently" into a catastrophe. These outcomes are all familiar, well-documented phenomena.

3.3.3.2 Definition

In complex and dynamic contexts, task requirements can emerge inconsistently, unpredictably, or even imperceptibly as the situation changes. The best way to perform a task can vary dramatically from one situation to the next, and the performance outcome may be dependent upon many interrelated variables, not all of which are under one's control. Changes in the overall context may generate entirely new operational requirements, while small but significant details may dictate the best way to perform a task. These factors and uncertainties demand constant attentiveness if one is to perform well, making the maintenance of good SA a major part of the job.

The concept of *situational awareness* first came to prominence in the aviation field, where enhanced SA in fighter pilots was found necessary to enable them to perform increasingly complex combat operations, while the loss of pilot SA was often found to be a factor in both civil and military accidents. The development of this concept has coincided with the ability to present substantially more data to pilots via their digital cockpit displays.

Endsley, like other human factors researchers, has emphasized that SA involves "far more than merely being aware of numerous pieces of data." It also requires a deeper understanding of the present situation as a whole and the implications for what is to come (Endsley, 1987).

3.3.3.3 Levels of Situational Awareness

With perception, the operator achieves a basic awareness of the concrete, objective elements that make up the operational environment: objects, events, persons, actions and so on, some of which may play a part in some situation or other. Perceptual alertness to one's environment is widely recognized as being merely one aspect or level of SA, which Endsley (1995a) has described as "Level 1". There are also higher (or deeper) forms of SA that rely upon the intelligent integration of current perceptual information with expert knowledge and cognitive skill to maintain both a coherent sense of meaning and a practical insight into the implications of the information.

Comprehension, Endsley's "Level 2" process in SA, means "going beyond the information given" by calling upon existing knowledge structures (schemas) to give meaning to perceptual experiences and perceived information. A pilot, for example, will routinely sample objective information from a flight display (e.g., "ALTITUDE: 2000 ft"), and rapidly translate the raw values into meaningful relations or states defined by the task objectives (e.g., "I'm slightly too high for this stage of final approach.").

Schema theory provides a plausible description of SA comprehension as the interpretation of current perceptions in the light of pre-established knowledge and expectations. The activation of knowledge (as a richly interconnected web of schemas) makes meaningful comprehension possible (Rumelhart & Norman, 1981). Just as expert readers develop meaning by applying knowledge that is not given in a text, so expert operators can seek to comprehend a situation by applying their specialist knowledge to information and intelligence obtained about the situation at hand. In highly familiar situations this process will occur automatically: the obvious meaning will quickly "leap out" to awareness. In novel situations, however, the operator must divert attentional time and resources to construct and test hypothetical accounts in order to make sense of the anomalous information.

3.3.3.3.1 Extending Situational Awareness

The processes of perception and comprehension are used in the *assessment* of the given situation: detecting things that are unexpected or potentially significant, interpreting what it all means, determining the severity of the situation, and so on. In parallel with this ongoing situation assessment, it is also important to be aware of the implications of these factors for the *management* of the situation: to understand what *could* happen in the near future with respect to one's objectives, and to determine what decisions *should* be made to satisfy those objectives (i.e., what actions one may take in order to succeed). The situation an operator faces is not just something that happens *to* the individual – if the operator needs SA, it is as an active participant *within* the situation, deliberately seeking to influence it or even control the situation.

Endsley refers to the first inferential process (inferring what is likely to happen) as *projection*, although it is not clear if this term is also explicitly intended to include what courses of action are available and likely to succeed. This is quite a different type of awareness. It is one thing to *project* that one's aircraft will crash if it maintains its present course; while it is quite another to *resolve* that the optimum way to avoid crashing is to turn and climb. Indeed, as our earlier quotation stated, the whole point of having SA is "to figure out what to do." Operators must act and must therefore be aware of how to act appropriately for the given situation, both now and in the future. For this reason, one can argue a case for extending Endsley's three-part model of SA by making this action-oriented process explicit as a fourth component. Hence, we now have four core elements of information processing and inferential reasoning, suggesting a four-way model of situational awareness as shown in Figure 5.

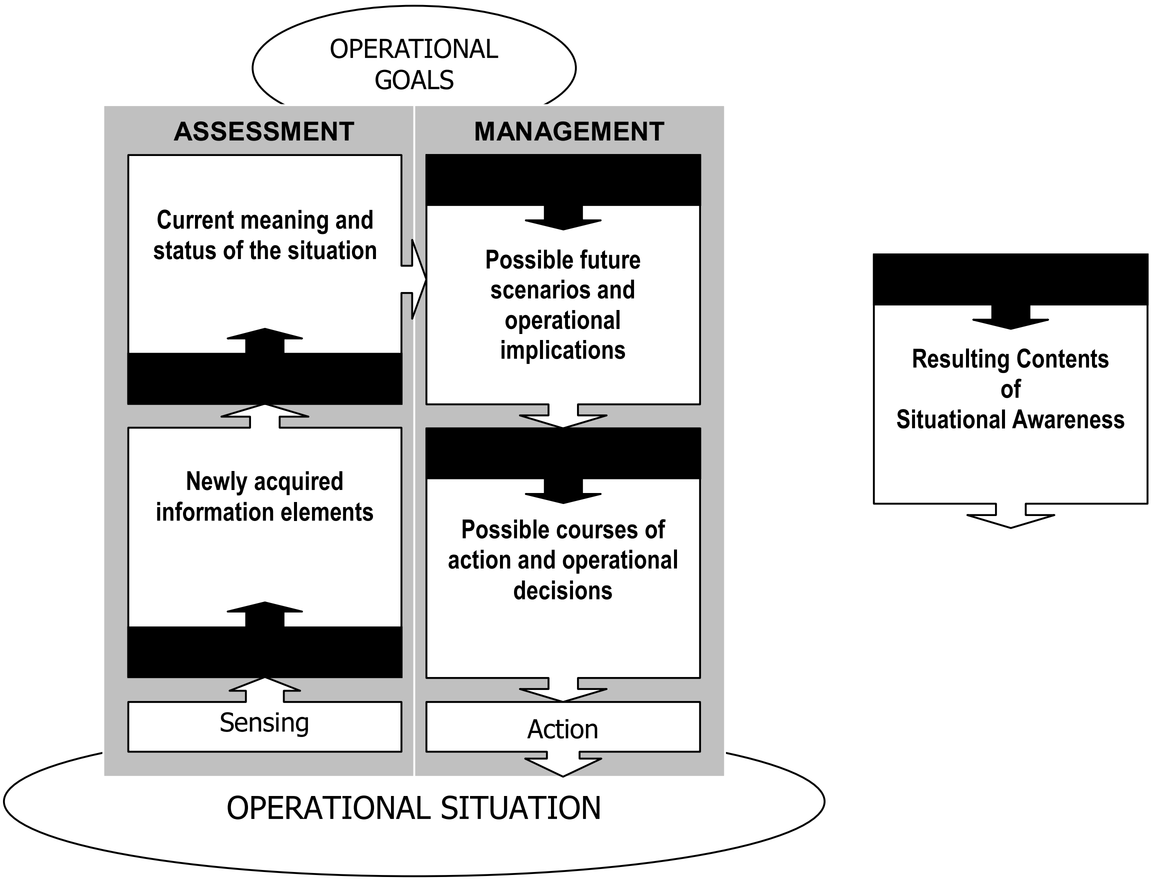

Figure 5: A Model of the Contents of SA and the Related Processes in the Assessment and Management of Operational Situations.

3.3.3.4 Measurement of SA

When it comes to measuring a person's (or team's) situational awareness, whether for the purposes of research and development, candidate selection, or training assessment, one can choose to focus on different aspects of SA:

- the accuracy and completeness of the *contents* of awareness (i.e., the validity of the person's inner model of operational reality),
- the performance of SA-related cognitive *processes* (perception, comprehension, etc.),
- the operational effects of SA on task *performance*, such as the time taken to respond to a significant change in the situation,
- the performance of essential *communications* or other information transactions in the sharing of SA in teams, and
- the recording of indirect *correlates* of SA, such as physiological indices and behavioral markers of different SA states or processes.

In addition, as with other cognitive phenomena such as mental workload, one can attempt to assess SA either *objectively* or *subjectively*. Objective measures include task performance and physiological correlates. Subjective measures include either self-ratings or expert observer-ratings.

3.3.3.4.1 Objective Techniques

Assessing the contents of SA seems like a peculiar mix of both objective and subjective approaches: one often *probes* the person's own "subjective" awareness of the situation, but then treats the responses as objective evidence of SA content, which can then be compared against the "ground truth" – the real situation itself. This technique has been embodied in several tools, most notably SAGAT (Situation Awareness Global Assessment Technique), a question-and-answer method developed by Endsley (1995b). In this case, a simulation exercise is interrupted and "frozen" while the subject is presented with several predetermined multiple-choice questions about the current situation (including its future developments). For example:

Q: What is the status of bridge Alpha?

1: Destroyed.

2: Standing and held by friendly forces.

3: Standing but held by enemy forces.

4: Standing, not yet held by either side.

When interruption is not appropriate or simulation freezing is not possible, "real-time" techniques may be considered. An example is the periodic request for a simple situation report or *Sit Rep*, specially formatted so as to probe the subject's awareness of key elements of the situation – including those that, ideally, they "should" be aware of – but without inadvertently alerting them to items they may have missed. For instance, one can provide section headings such as "Latest known enemy movements." This is a relatively non-intrusive technique and holds a certain familiarity for most subjects who are accustomed to giving situation reports.

Another technique is similar to what is known in psychology as the Sentence Verification Task, whereby the subject is presented with a series of descriptions of the situation (e.g., "Enemy patrol sighted approaching bridge Alpha"), and is asked to assess the veracity of each statement, rating it as either True or False. A method for statistically analyzing verification responses, which is similar to that used in Signal Detection Theory, is being developed at BAE SYSTEMS under the name of QUASA (Quantitative Assessment of Situational Awareness).

There are two essential points to make with respect to objective, probe-type assessments of SA content:

1. In order to prepare appropriate probes that are meaningful and relevant to the subject, the researcher must take the time to analyze and understand the subject's task-specific awareness needs prior to the study.
2. In order to asses the subject's responses, the researcher needs a record of the actual situation at that time – not merely the objective composition but also the operational meaning and implications.

For both of these requirements, the assistance of a subject-matter expert is usually essential.

3.3.3.4.2 Subjective Techniques

Subjective ratings of SA can be given either by an expert observer (observer ratings) or by the operator (self ratings). The rating instruments are normally designed to elicit ratings of either SA content (e.g., the validity of the subject's understanding of the situation) or SA-related processes (e.g., the effectiveness of the subject's information monitoring). They can consist of a single ("unidimensional") rating scale, such as a modified version of the Bedford Scale used for workload rating, or several different ("multidimensional") scales. Unidimensional scales are generally quick and easy to administer, while multidimensional scales are designed to tease apart different aspects of SA and to evaluate them separately.

SART (Situation Awareness Rating Technique) is a multidimensional self-ratings technique developed by Taylor and Selcon (1991). It consists of a set of ten self-rating scales presented to the operator after a trial or simulation run. The ten scales represent ten "core dimensions" of SA that emerged from a Repertory Grid analysis of a number of pilots' "personal constructs" of SA. The dimensions are labeled as follows:

1.	*Instability*	Likeliness of situation to change suddenly.
2.	*Variability*	Number of variables that require one's attention.
3.	*Complexity*	Degree of complication (number of closely connected parts) of situation.
4.	*Arousal*	Degree to which one is ready for activity.
5.	*Spare capacity*	Amount of mental ability available to apply to new variables.
6.	*Concentration*	Degree to which one's thoughts are brought to bear on the situation.
7.	*Division of attention*	Degree of distribution or focusing of one's perceptive abilities.
8.	*Information quantity*	Amount of knowledge received and understood.
9.	*Information quality*	Degree of goodness or value of knowledge communicated.
10.	*Familiarity*	Degree of acquaintance with situation through experience.

On close inspection, it can be seen that most of these constructs refer to things related to SA and its acquisition, but not to situational awareness itself. In other words, SART does not reveal a person's level of SA, either actual or perceived. On the other hand, the provided ratings may be highly relevant to the objectives of the trial and may give vital insights into the different variables affecting SA.

CARS (Crew Awareness Rating Scale) is a self-ratings tool for the subjective assessment of SA in terms of both content and the related cognitive processing. Its purpose is to elicit an operator's personal evaluation of his/her experience of SA content and processes in terms of the four aspects referred to in the model above (Figure 5). Thus, there are eight rating scales:

1. *Perception* – contents
2. *Comprehension* – contents
3. *Projection* – contents
4. *Resolution* – contents
5. *Perception* – processing
6. *Comprehension* – processing
7. *Projection* – processing
8. *Resolution* – processing.

The elicited ratings data hopefully complement simultaneous objective assessments (see next section). For each question, the subject is asked to rate the extent to which he/she *has* good SA (content) or *is able to maintain* good SA (processing). The subject is also given the option of responding "Don't know," in effect demonstrating an absence of self-perception for that aspect.

3.3.3.5 Combining Objective Measures with Self-Ratings

Self-ratings of SA content provide the individual's assessment of awareness. Some researchers refer to this method as "perceived SA" to contrast it with the "actual SA." This subjective approach does not, of course, provide a true measure of SA content. Rather, self-ratings reflect only the self-perception – the subjective assessment of SA. For example, a positive self-assessment indicates high confidence in the individual's SA, while a negative self-assessment indicates low confidence in SA, or a belief that the operator is disoriented, confused, under-informed or "losing the plot."

Subjective evaluations of oneself are notoriously prone to distortion. For example, one's actual SA could be quite good, yet for some reason one might rate it as poor. However, this self-assessment, distorted or not, *is itself a component of SA* for the simple reason that *the person himself, including his/her SA, is a component of the operational situation*. In a circular fashion, SA can be affected positively or negatively by how positively or negatively the individual assesses its content. Thus, self-perception or perceived SA interacts with actual SA, and this interaction has an influence on decision-making and confidence in pursuing a course of action. This notion is illustrated in Figure 6, where the four quadrants represent the different combinations of high/low actual SA versus positive or negative self-assessment of SA (McGuinness, 1995).

		ACTUAL SA	
		LOW	**HIGH**
SELF-ASSESSMENT	**POSITIVE**	**WORST CASE:** Inappropriate Confidence	**IDEAL CASE:** Appropriate Confidence
	NEGATIVE	Appropriate Caution	Inappropriate Caution

Figure 6: Combinations of Actual SA (low or high) versus Self-Assessment of own SA (positive or negative).

In reality, the levels of actual SA and SA self-assessment are more likely to be on continua, but the matrix serves to simplify the differences between combinations and reveals four extreme cases:

1. WORST CASE: The individual has low SA, but perceives it as high. The operator has inappropriate confidence in his/her own SA, and thus makes poor decisions with inappropriate confidence. This is where decision errors leading to performance failures or accidents are most likely to occur.
2. APPROPRIATE CAUTION: The individual has low SA, and perceives it correctly. He/she is appropriately unconfident in the SA and is suitably cautious in decision-making while trying to regain high SA.
3. INAPPROPRIATE CAUTION: The individual has high SA, but for some reason perceives it as low. The person is under-confident and is inappropriately cautious in decision-making.
4. IDEAL CASE: The individual has high SA, and perceives it correctly. The operator has appropriate confidence and is capable of making good decisions.

This analysis of different combinations of actual and perceived SA demonstrates why self-ratings are extremely useful. If one assesses only the actual content of the individual's awareness, there is no identification of whether the self-assessment of awareness is positive or negative. At the same time, one should take care to not rely solely on self-ratings without comparing them with a more objective assessment of actual SA. If one records only self-assessments of SA, there is no confirmation of whether the actual awareness is high or low. To reiterate: both the actual SA and the self-assessment of SA (perceived SA) have a significant bearing on the individual's decision-making and subsequent performance. It is therefore strongly recommended that, when possible, self-ratings be taken in conjunction with more objective measure of actual SA such as content probes.

3.3.3.6 References

Adam, E.C. (1993). Fighter Cockpits of the Future. In *Proceedings of the 12th DASC, the 1993 IEEE/AIAA Digital Avionics Systems Conference* (pp. 318-323). New Jersey: Institute of Electrical and Electronic Engineers.

Endsley, M.R. (1987). *SAGAT: A Methodology for the Measurement of Situation Awareness* (In Report No. NORDOC 89-32). Hawthorne, CA: Northrop Corporation.

Endsley, M.R. (1990). A Methodology for the Objective Measurement of Pilot Situation Awareness. In *Situation Awareness in Aerospace Operations* (AGARD-CP-478, pp. 1-1–1-9). Neuilly-sur-Seine, France: NATO Advisory Group for Aerospace Research and Development.

Endsley, M.R. (1995a). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, *37*, 32-64.

Endsley, M.R. (1995b). Measurement of Situation Awareness in Dynamic Systems. *Human Factors*, *37*, 65-84.

McGuinness, B. (1995, November). *Situational Awareness Measurement in Human Factors Research* (BAE SRC Report JS13292). Bristol, England: BAE Systems.

Rumelhart, D.E., & Norman, D.A. (1981). *Analogical Processes in Learning* (pp. 335-361). In J.R. Anderson (Ed.), Cognitive Skills and their Acquisition. Hillsdale, NJ; Erlbaum.

Taylor, R.M., & Selcon, S.J. (1991). Subjective Measurement of Situational Awareness. In *Proceedings of the 11th Congress of the International Ergonomics Association* (pp. 789-794). Paris: IEA.

Chapter 4 – ASSESSMENT METHODS

4.1 PHYSIOLOGICAL MEASURES

4.1.1 Actigraphy

4.1.1.1 Description of Actigraphy

Wrist-mounted actigraphy was developed in the 1970s and 1980s. The wristwatch-sized unit contains accelerometers that respond to arm movements. If the magnitude of a movement exceeds a preset threshold then an event is registered. The number of events occurring within a pre-selected time interval is stored. This event count provides a measure of limb activity over time. For example, activity can be recorded in one-minute intervals continuously across hours and days.

4.1.1.2 Background

Actigraphy was originally developed to objectively measure and quantify sleep based on body movements, prior to the development of polysomnographic techniques. The first study involving actigraphy was performed by Szymansky (1922), who constructed a device that was sensitive to the gross body movements of subjects as they lay in bed. However, the advent of EEG recording techniques and their application to sleep (Loomis, Harvey, & Hobart, 1937), followed by the institution of EEG-based standards for the scoring of sleep stages (Rechtschaffen & Kales, 1968), caused a shift in interest away from movement-based measurements of sleep.

The development of wrist-mounted actigraphy generated a resurgence of interest in the movement-based measurement of sleep. This interest also was fuelled by technological advances that, for the first time, made portable measurement and recording of movement data over long periods (days, weeks, or even months) feasible. Furthermore, even with portable ambulatory EEG recorders, EEG-based measurement of sleep and wakefulness was neither logistically practicable nor cost-effective for determining basic sleep/wake rhythms in large numbers of subjects and/or when the study period of interest lasted several weeks or months.

With the development of technologically advanced actigraph components, the primary issue became the extent to which actigraphic measures of sleep/wake state were both reliable and valid compared to the gold standard of polysomnography (PSG) for recording sleep/wake periods. Several validation studies have subsequently been performed using different actigraph scoring algorithms, subjects from various age ranges, varying sample sizes, and subjects with various sleep and/or movement-related disorders. These studies are reviewed below. For a recent review and discussion of clinical issues, see Sadeh, Hauri, Kripke, and Lavie (1995). In general, such studies indicate that wrist actigraphy is a valid and objective measure of sleep/wake state (Sadeh et al., 1995).

An early pilot study to address validation issues was conducted by Kripke, Mullaney, Messin, and Wyborney (1978). Using five normal subjects, they reported excellent agreement between actigraphically-derived, manually scored measures and PSG-determined, manually scored measures of sleep duration. Kripke et al. (1978) reported a correlation coefficient of 0.98, a correlation higher than a typical correlation between two well-trained individuals manually scoring a PSG (which is generally in the 0.90 range). Shortly thereafter, the same research group published results from a larger-scale validation study in which actigraphically-determined and polysomnographically-determined sleep/wake estimates were compared from a total of 102 nights. This study included data from 39 hospital patients and 63 non-patients (Mullaney, Kripke, & Messin, 1980). Overall, the two methods produced an agreement rate of 94.5% (i.e., 94.5% of the one-minute epochs were manually scored correctly using actigraphic methods, using "blind" manual PSG scoring as the "gold standard"). When the sub-sample of hospital patients was excluded from the analyses, the agreement rate rose to 96.3%. Significant correlations were obtained in this study for a number of manually scored sleep parameters including Total Sleep Time (TST; $r = 0.89$) and minutes of Wake time After Sleep Onset (WASO; $r = 0.70$). Not all actigraphically-determined sleep parameters were significantly correlated with their polysomnographically-determined counterparts. For example, actigraphy proved relatively poor for specifying the actual number of discrete mid-sleep awakening events ($r = 0.25$).

Using college students as subjects (n = 14), Webster, Kripke, Messin, Mullaney, & Wyborney (1982) reported an overall agreement rate of 93.9% between PSG and actigraphic measures of sleep/wake. This study differed from those reported above in that the actigraphic records were scored automatically using a sleep/wake scoring algorithm. Thus, Webster et al. (1982) also published the first algorithm that could be used to automatically score actigraphic data, an important step since up to that point the labor-intensive and tedious task of manually scoring actigraphic data on an epoch-by-epoch basis at least partially offset the advantages of the data collection technique.

4.1.1.3 State of the Art

Recently developed actigraph units provide on-line analysis of activity data, which extends the capabilities of earlier units that only detected motion and stored the data. The Sleep Watch Actigraph (SWA – see Figure 7) contains a central processing unit, random access memory, and an accelerometer. Each minute, the SWA records whether and how much movement activity has occurred. If acceleration of the wrist changes, the accelerometer generates a small electrical current. If the electric current exceeds a specified threshold, a "1" is recorded; otherwise, a "0" is recorded. The "1" or "0" is stored in the device. In this way, activity is recorded in one-minute intervals continuously across hours and days.

Figure 7: Sleep Watch Actigraph showing Fuel Gauge-Type Current Performance Capacity Readout.

Built into the SWA is a sleep-scoring algorithm that takes the minute-by-minute activity score and determines if the wearer is awake or asleep. Also built into the SWA is the Sleep Performance Prediction Model (SPM) described in Balkin et al. (2000). The SPM takes the output of the sleep-scoring algorithm (the wearer's sleep/wake history) and uses this information to predict changes in performance in real time. The SPM includes a charging function for recuperation during sleep, a linear decline in performance while awake, and a circadian rhythm modulating function with the acrophase, or peak, set at 2000 hours. The SWA device has a display that includes both an analog and digital "fuel gauge." These gauges indicate the current SPM performance prediction. The digital gauge displays the wearer's performance prediction as a percentage of full capacity. The SWA device also includes a light sensor. Light is the primary determinant of the circadian rhythm of alertness (Duffy, Kronauer, & Czeisler, 1996). Future SWA models will include a function to adjust the circadian rhythm for time zone changes based on the actual history of bright light exposure.

4.1.1.4 Limitations

4.1.1.4.1 What Actigraphy Can Tell Us

In addition to its usefulness for measuring and recording sleep/wake history (from which estimates of performance capacity can be generated), there is considerable evidence that the wrist actigraph signal may be empirically useful in other ways. For instance, it appears that threshold count data taken from an actigraph set to pass signals within the 0.1 to 3 Hz bandwidth tend to settle during rest at a count at or near the heart rate (instead of zero, when a passband of 2-3 Hz is used). Indeed, it has been found that the sensor signal contains a very low-level ballistographic signature of the heartbeat, as well as a low-frequency variation suggestive of breathing movement, when not masked by larger amplitude movements (see Figure 8).

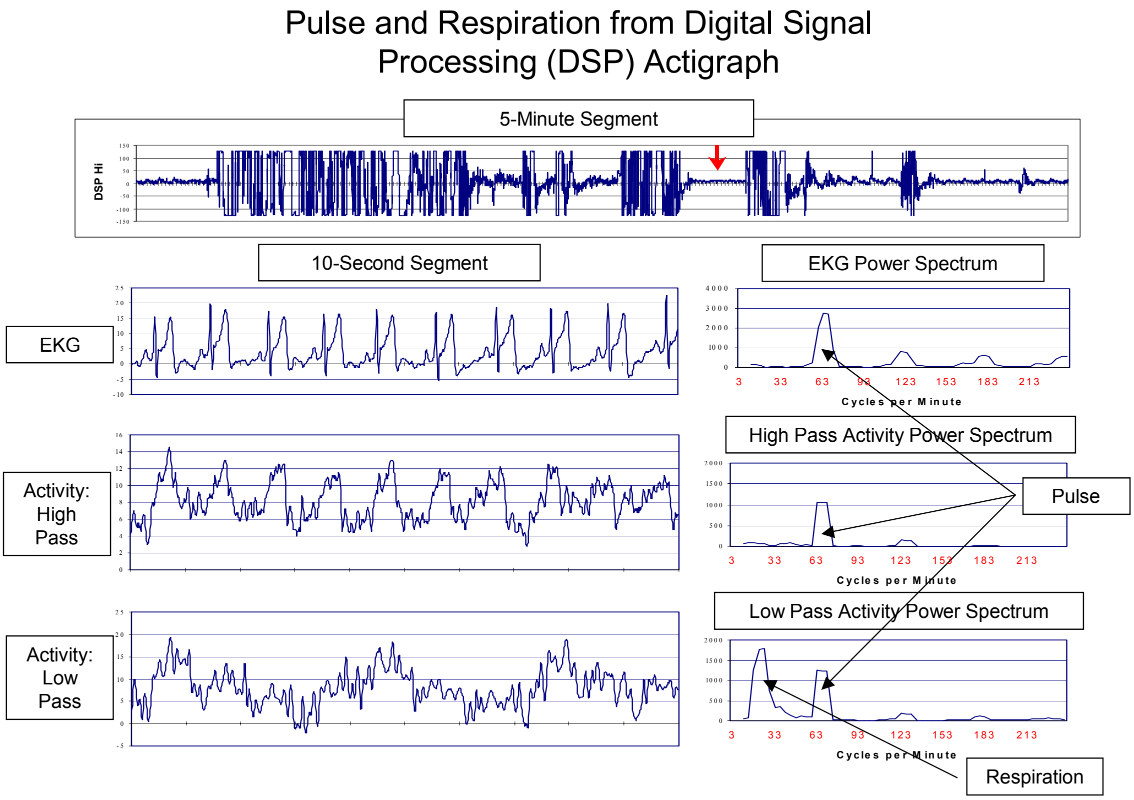

Figure 8: Heart and Respiration Rate Signals are Detectable with Proper Filtering of Actigraph Data.

Furthermore, when the passband is set to the full range of 0.1 to 9 Hz and sensitivity is maximized, the actigraph registers non-zero counts continuously, as long as the device is being worn. Precision Control Design, Inc (maker of the AMA-32 actigraph) exploits this phenomenon as “LifeSign” data, using it to detect when the actigraph is off the wrist. The source of this data stream is uncertain and warrants further investigation since it appears to be biological in origin. It may be related to “microvibrations,” which were described by Rohracher (1960) but were never fully examined or put to a useful purpose. According to Rohracher, this low-level tremor occurs in the frequency band of 7.5 to 12.5 Hz, and so would be readily detected by the actigraph sensor at broad passband settings. While "outside the envelope" of standard actigraphy, the questions of whether extraction of heart rate, breathing rate, and microtremor is possible by this method, and whether the information may be useful in discriminating sleep stages or sleep stage transitions should be evaluated.

4.1.1.4.2 What Actigraphy Cannot Tell Us

The level of sleep debt can be determined from wrist actigraphy, and this information, in combination with circadian rhythm information, can be used to predict performance capacity. However, actigraphy does not measure nor predict moment-to-moment fluctuations in alertness or performance capacity. At best, it defines the likely *range* within which performance and alertness will vary on a moment-to-moment basis.

Also, although wrist actigraphy provides an accurate measure of the sleep/wake schedule of its wearer, the embedded Sleep Performance Model (SPM) is not yet individualized. That is, the SPM predicts the effect that the wearer's sleep schedule would have on an average person, but it does not yet take into account individual differences in sleep need or performance capacity. Future versions will include the capacity for the Sleep Watch to *learn* the extent to which the wearer's variations in sleep schedule affect that individual's performance.

Conventional actigraphic design represents an optimization of past technology based on two key considerations: (1) consistent reliability of the output data (counts of threshold crossings) as input for the detection of sleep/wake state transitions using validated weighted moving average algorithms such as that of Cole, Kripke, Gruen, Mullaney, & Gillin (1992), and (2) size, weight, power requirement, and other electrical and electronic features realizable as a user-accepted device of reasonable cost. Currently, this optimization produces very sharp and deliberate limitations of the information originally contained in the movement signal and passed on to the scoring algorithm. As discussed in Redmond and Hegge (1985), there are four main areas of design constraint:

1. The sensitivity of the sensor must be such as to respond to "normal" arm movements, but not be "swamped" by the waking movements of a very active person, or by sources of external noise and vibration. Information from very fine, subtle movement is sacrificed.
2. The frequency response of the accelerometric sensor system is sharply confined to a frequency band of 2 to 3 cycles per second (Hz). At the low end, this filtering is needed to eliminate counts from undulating, slow-wave excursions of the sensor (e.g., due to breathing, rocking of the device in the gravitational field, or vehicle motion) that are not actually due to motor activity. At frequencies above 3 Hz, this response helps eliminate false counts due to tremor, external noise and vibration, and "ringing" due to sharp impulses.
3. The translation of a complex movement signal into a simple measure, readily computed and expressed digitally in microprocessors of 1985-1995 vintage, resulted in the use of threshold-crossing counts, but eliminated far more descriptive measures of the signal characteristics, such as duration, amplitude, and power.
4. The use of extended periods of measure relative to movement rates (i.e., 1- or 2-minute bins) keeps data sets at a workable length in electronic memory, and matches the temporal scale expected by validated sleep/wake algorithms. The integration of sensor data over time smoothes over transient bursts of sensor activity. This smoothing may or may not be advantageous depending on whether such transients are themselves physiologically relevant.

Recognizing that usage of the existing actigraph thus filtered out a large portion of information contained in the original raw movement signal, the actigraph was redesigned to permit the automated setting of alternate sensitivities, counting thresholds, and frequency response bands. The design intent was to allow investigation of varied settings (or information content), while normal usage emulated the original, standardized settings of "High Gain, High Threshold," and 2 to 3 Hz bandwidth. In 1993, Elsmore and Naitoh compared the varied actigraph settings against PSG-scored sleep using three actigraph/sleep algorithms (Cole et al., 1992; Sadeh, Alster, Urbach, & Lavie, 1989, and Pleban, Valentine, Penetar, Redmond, & Belenky, 1990). Their report confirmed agreement with PSG sleep in the range of 79% to 93% for standard actigraph settings, using both the Cole and Sadeh algorithms. However, the authors found that the broad-band frequency settings (0.1 to 3 or 9 Hz) and the low threshold setting produced such high counts in sleep as to render the standard algorithms useless.

The experience above and others described by Elsmore, as well as those at Walter Reed, point again to a fundamental limitation when using the actigraph to explore *outside* the bounds of optimization. The chosen settings for gain, threshold, and passband are arbitrary (albeit grounded in the original studies of Redmond & Hegge, 1985), with no means of readily adjusting them for comparison's sake while controlling for movement events (system input). Selection of a particular combination of passband, gain, threshold, and digital counting transform automatically selects out other features of the signal's complexity, potentially *distorting* the original information contained in it, as reported at the output. A systematic approach to this problem requires continuous access to the raw unfiltered signal, and the computational means for parsing, manipulating, and statistically treating its information content.

Actigraphy has been found to have low correlation with heart rate measures in nurses and healthy elderly subjects during their normal daily activities (Goldstein, Shapiro, Chicz-DeMet, & Guthrie, 1999; Shapiro & Goldstein, 1998). Thus, actigraphic data cannot currently be used to determine the effects of cognitive and physical activity on an operator's cardiovascular system.

In short, definitive treatment of wrist-movement characteristics vis-à-vis sleep-related events, and the subsequent design of actigraphic devices capable of more than simple sleep/wake discrimination, await (1) systematic study of the fundamental contents of the sensor-signal driven by movement behavior in both sleep and waking states, and (2) enabling technology for conducting such research and device development.

4.1.1.5 General Advantages/Disadvantages of Actigraphy

Wrist actigraphy is non-intrusive to the wearer. Since current versions (like the Sleep Watch, see Figure 7) have all the functions of a typical sports-style wristwatch, the wearer can seamlessly substitute one of these units for his or her own wristwatch. The device is easily programmed by technicians and the downloading and analysis of the data are automated. While not currently available, it will be possible to obtain information about entire military units by combining actigraph data from the individual operators.

4.1.1.6 Apparatus Required

An actigraph recorder is required for each operator. Additionally, a PC computer is required for programming and data downloading. One such computer can be used to collect and analyze the data from a large number of individual actigraph recorders. Special software is required for programming the recorders and analyzing the data. This software is commercially available.

4.1.1.7 Personnel Required

Personnel are needed to maintain the actigraph recording units. They need to test the recorders to assure that they are operating correctly. Batteries must be checked and replaced. Current battery life ranges from 6 to 8 weeks depending on actigraph sampling rate, but it is anticipated that batteries lasting 6 months to 1 year will be available by 2005. Data downloading and analysis do not require high skill levels since these functions are automated, and current technology will allow remote, telemetered downloading for automated scoring of data.

4.1.1.8 Analysis Techniques

Automatic scoring algorithms are used to analyze the data for general activity levels, for circadian rhythms, and for sleep maintenance analysis. These procedures take advantage of the extensive studies that have been conducted using actigraph technology. Although several scoring algorithms have been developed, the Cole-Kripke algorithm (Cole et al., 1992) has undergone the most extensive testing and validation, and is currently the most widely used in both clinical and operational environments.

4.1.1.9 References

Balkin, T., Thorne, D., Sing, H., Thomas, M., Redmond, D., Wesensten, N., Williams, J., Hall, S., & Belenky, G. (2000). *Effects of Sleep Schedules on Commercial Driver Performance* (Report No. DOT-MC-00-133). Washington, DC: U.S. Department of Transportation, Federal Motor Carrier Safety Administration.

Cole, R.J., Kripke, D.F., Gruen, W., Mullaney, D.J., & Gillin, J.C. (1992). Automatic Sleep/Wake Identification from Wrist Actigraphy. *Sleep*, *15*, 461-469.

Duffy, J.F., Kronauer, R.E., & Czeisler, C.A. (1996). Phase-Shifting Human Circadian Rhythms: Influence of Sleep Timing, Social Contact, and Light Exposure. *Journal of Physiology*, *495*, 289-297.

Elsmore, T.F., & Naitoh, P. (1993). *Monitoring Activity with a Wrist-Worn Actigraph: Effects of Amplifier Passband and Threshold Variations* (NHRC Tech Report 93-18). San Diego, CA: Naval Health Research Center.

Goldstein, I.B., Shapiro, D., Chicz-DeMet, A., & Guthrie, D. (1999). Ambulatory Blood Pressure, Heart Rate, and Neuroendocrine Responses in Women Nurses during Work and Off Work Days. *Psychosomatic Medicine*, *61*, 387-396.

Kripke, D.F., Mullaney, D.J., Messin, S., & Wyborney, V.G. (1978). Wrist Actigraphic Measures of Sleep and Rhythms. *Electroencephalography and Clinical Neurophysiology*, *44*, 674-676.

Loomis, A.L., Harvey, E.N., & Hobart, G.A. (1937). Cerebral States during Sleep, as Studied by Human Brain Potentials. *Journal of Experimental Psychology*, *21*, 127-144.

Mullaney, D.J., Kripke, D.F., & Messin, S. (1980). Wrist-Actigraphic Estimation of Sleep Time. *Sleep*, *3*, 83-92.

Pleban, R.J., Valentine, P.J., Penetar, D.M., Redmond, D.P., & Belenky, G. (1990). Characterization of Sleep and Body Composition Changes during Ranger Training. *Military Psychology*, *2*, 145-156.

Rechtschaffen, A., & Kales, A. (Eds.). (1968). *A Manual of Standardized Terminology, Techniques, and Scoring System for Human Subjects* (NIH Publication No. 204). Washington, DC: U.S. Government Printing Office.

Redmond, D.P., & Hegge, F.W. (1985). Observations on the Design and Specifications of a Wrist-Worn Human-Activity Monitoring System. *Behavioral Research Methods, Instruments, & Computers*, *17*, 659-669.

Rohracher, H. (1960). Methods for Registration and Evaluation of Micro Vibration. *Psychologische Beitraege*, *4*, 118-126.

Sadeh, A., Alster, D., Urbach, D., & Lavie, P. (1989). Actigraphically Based Automatic Bedtime Sleep/Wake Scoring: Validity and Clinical Applications. *Journal of Ambulatory Monitoring*, *2*, 209-216.

Sadeh, A., Hauri, P.J., Kripke, D.F., & Lavie, P. (1995). The Role of Actigraphy in the Evaluation of Sleep Disorders. *Sleep*, *18*, 288-302.

Shapiro, D., & Goldstein, I.B. (1998). Wrist Actigraph Measures of Physical Activity Level and Ambulatory Blood Pressure in Healthy Elderly Persons. *Psychophysiology*, *35*, 305-312.

Szymansky, J.S. (1922). Aktivitaet und ruhe bei den menschen. *Z Angew Psychology*, *20*, 192-222.

Webster, J.B., Kripke, D.F., Messin, S., Mullaney, D.J., & Wyborney, G. (1982). An Activity-Based Sleep Monitor System for Ambulatory Use. *Sleep*, *5*, 189-199.

4.1.2 Cardiorespiratory Measures

4.1.2.1 Description of Cardiorespiratory Measures

Cardiorespiratory measures involve measures of the cardiac and respiratory systems such as heart rate, blood pressure, respiratory frequency, respiratory amplitude, and oxygen consumption. The cardiovascular system transports blood to all organs of the body. The pump output of the heart is altered by a beat-by-beat adjustment in its rate and force. Such a constant regulation of the cardiovascular system provides the physiological basis for heart rate, heart rate variability, and blood pressure. The mechanism of cardiovascular regulation is complex. There are not only multiple control mechanisms, but also many complex feedback loops to regulate various parts of the cardiovascular system and their interactions. Heart rate variability has become a popular state assessment technique because it can be obtained relatively easily and provides information about changes in the cardiovascular control system.

4.1.2.2 Background

Cardiovascular and respiratory measures have been used for operator functional state (OFS) assessment for many years. In the nineteenth century, measuring heart rate was a standard tool for assessing the state of patients. Systematic changes in heart rate were found as a function of relaxed vs. excited states, low vs. high body temperatures, and between levels of exercise. In the second half of the nineteenth century, systematic relationships among various cardiovascular and respiratory measures were identified. For example, it was found that heart rate increases during inhalation and decreases during exhalation. In 1876, Mayer described systematic fluctuations in blood pressure occurring at frequencies of 6 to 9 cycles/min (0.1 to 0.15 Hz), which correspond to heart rate variations in the same frequency range and are still referred to as Mayer waves in the current literature.

The measurement of cardiac activity has been a popular physiological technique for the assessment of mental effort and workload during the last three decades (Wierwille & Eggemeier, 1993). The sensitivity of different cardiac measures (electrocardiogram-ECG, blood pressure, and blood volume) to variations in mental workload has been examined extensively. Heart rate (HR) and HR variability (HRV) have been the most promising measures (Heart Rate Variability, 1996; Hopman, Kollee, Stoelinga, van Geijn, & van Ravenswaaij-Arts, 1993; Kramer, 1991). Many advanced mathematical methods have been introduced to analyze the dynamics of heart rate.

In the twentieth century, systematic changes in heart rate and heart rate variability due to stress were identified. In 1963, Kalsbeek and Ettema suggested that changes in the variability of the instantaneous heart rate could be used to indicate changes in mental load. This observation stimulated extensive investigation of heart rate variability and the assessment of mental workload.

Measuring cardiovascular parameters without knowledge of the underlying control system does not provide adequate information about operator state. Therefore, a brief outline of the cardiovascular control system is provided. Figure 9 presents a simplified model. The cardiovascular control system can regulate blood pressure (BP) by adjusting venous volume, vascular resistance, contraction force, and heart rate. For example, when BP decreases, an increase in HR will be found almost immediately. The main neural mechanisms for regulating BP are the sympathetic and the parasympathetic branches of the autonomic nervous system. An increase in sympathetic activity results in an increase in HR, and an increase in parasympathetic activity results in a decrease in HR. BP changes are primarily the result of changes in the sympathetic system.

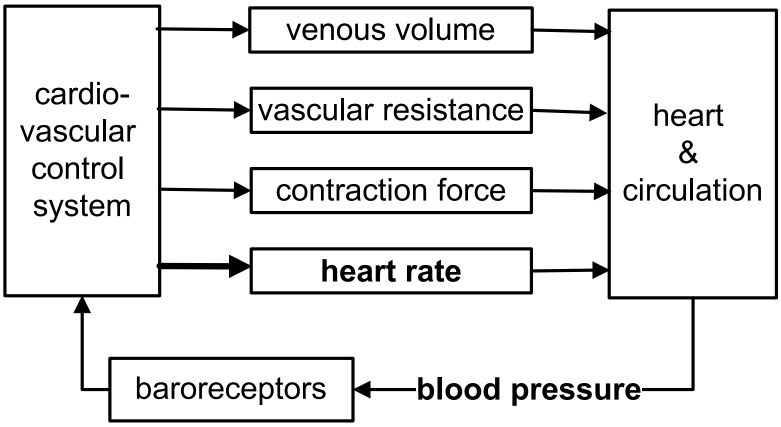

Figure 9: Simplified Model of the Baroreflex Loop.

Baroreceptor sensitivity, which indicates the flexibility of the physiological system to adapt to changes, can provide valuable information about OFS. For example, when operators invest substantial mental effort to cope with task demands, changes in BP are less reflected in changes in HR. This is an important cause of the reduction in heart rate variability (HRV) often found for operators during mental effort investment. A reduction in baroreceptor sensitivity is indicated by a reduced gain between HRV and BP variability (BPV). This combined approach provides a better indication of functional state than using only HRV. The gain between BPV and HRV has been found sensitive to mental effort (Veltman & Gaillard, 1996).

A common approach for measuring HRV is to use frequency spectra, which are often divided into several ranges. The European Society of Cardiology and the North American Society of Pacing and Electrophysiology (Malik et al., 1996) use four ranges: ultra low frequency component (ULFC; <= 0.003 Hz), very low frequency component (VLFC; 0.003-0.04 Hz), low frequency component (LFC; 0.04-0.15 Hz) and high frequency component (HFC; 0.15-0.40 Hz). In psychophysiological research, the frequency spectrum is often divided into components covering different ranges. For instance, Mulder (1988) uses three ranges: low-band (0.02-0.07 Hz), mid-band (0.07-0.15 Hz), and high-band (0.15-0.50 Hz). For the assessment of functional state during task performance (operational state), the exact definition of the various ranges is less important. The most important ranges are those around 0.10 Hz and around the respiratory frequency (about 0.3 Hz). These frequencies are distinguished in both definitions of frequency bands described above.

The frequency ranges reflect different mechanisms. The high-frequency band incorporates respiratory activity. During inhalation, HR increases and during exhalation, HR decreases, resulting in HR changes with the same frequency as the respiratory frequency (normally about 0.3 Hz). For this reason, the high band is often called "respiratory sinus arrhythmia". The cardiovascular control system is a closed-loop system with a resonance frequency of about 0.10 Hz, associated with changes in BP at this frequency. These changes are reflected in changes in HR, producing a peak in the HR spectrum around 0.10 Hz. For this reason, the frequency range around 0.10 Hz is often called the "0.10 Hz component" or the "blood pressure component". The regulation of body temperature causes changes in blood pressure that are reflected in the low band. Therefore, the low band is often referred to as the "temperature component".

Parasympathetic activity is the major contributor to the high-frequency component (above 0.15 Hz). Low-frequency oscillations are generally governed by sympathetic activity, although some studies show that the LFC reflects both sympathetic and parasympathetic activity. Consequently, the LFC/HFC ratio is considered by some investigators to mirror sympatho/vagal balance (or reduced sympathetic modulations). Physiological interpretation of VLFC and ULFC warrants further investigation. VLFC is typically associated with thermo-metabolic (humoral) modulations while ULFC reflects body movements and 24-h periodicity.

4.1.2.3 State of the Art

4.1.2.3.1 Heart Rate

Heart rate (HR) has been a popular psychophysiological measure to monitor operator state. The reason for this is that HR is easy to measure and has been proven sensitive to different operator states. HR increases due to both physical and mental activities. When the body requires more oxygen due to physical activity, the heart pumps more powerfully and HR increases. HR is also sensitive to mental effort. Numerous studies have found systematic relationships between cognitive demands and HR (e.g., Roscoe, 1992; Veltman & Gaillard, 1996, 1998; Caldwell et al., 1994).

4.1.2.3.2 Heart Rate Variability

The use of HRV in both laboratory and field settings is valued not only because of its usefulness as a measure of mental effort, but also in applications where continuous recording is required (Tattersall & Hockey, 1995). In laboratory studies, HRV has consistently responded to changes from rest to task conditions and to a range of between-task manipulations (Aasman, Mulder, & Mulder, 1987; Sirevaag et al., 1993). In operational contexts, HRV has seen increased use as an indicator of the extent of task engagement in information processing requiring significant mental effort, particularly in flight-related studies (Kramer, 1991; Sirevaag et al., 1993; Tattersall & Hockey, 1995; Wilson, 1993; Wilson & Eggemeier, 1991). HRV has been reported to respond rapidly to changes in operator workload and strategies, usually within seconds (Aasman et al., 1987; Coles & Sirevaag, 1987). Thus, HRV has been able to detect rapid transient shifts in mental workload (Kramer, 1991).

Aasman et al. (1987) found HRV to be associated with changing levels of user effort. In their study, participants were given simple (non-counting) and complex (counting) versions of a task. The study showed that the amplitude of the 0.10 Hz component of the cardiac interval signal was particularly affected in the complex task condition, as long as the subjects were working within the limits of working memory. When the limits of working memory were exceeded, most subjects were unable to cope with the demands of the task as evidenced by a performance decrement and an *increase* in HRV. Thus, when working memory was exceeded, participants gave up, indicating that less effort was invested.

In a study involving the level of user control and changes in HRV during simulated flight maintenance, the demands of dynamic monitoring and fault diagnosis for eleven trainee flight engineers were examined in relation to changes in HRV (Tattersall & Hockey, 1995). HRV was found sensitive to the different phases of the work task. In particular, the 0.07 - 0.14 Hz frequency range was suppressed during the mentally demanding problem solving phase. The findings of this study support both the use of HRV as a physiological index of mental effort and its value in operational contexts. The way in which HRV changes during mental and physical loading depends on the balance between LFC and HFC during baseline (Kepezenas & Zemaityte, 1983). When there is a slight prevalence of the high-frequency component during baseline, then mental effort will result in a decrease in HRV. When there is a strong prevalence of parasympathetic control (reduced HRV and very low HR frequency) during baseline (which can be found among well-trained sportsmen), then mild mental effort is often followed by an increase in HRV (both LCF and HFC). Only a high level of mental effort is characterized by a further decrease in HRV. Well-trained sportsmen are characterized by a low HR and decreased HRV (especially LCF) during baseline.

HRV might be used in OFS assessment during sleep-wake cycles. Studies of long-term acclimation of cardiac rhythm to microgravity in astronauts have shown that a more pronounced decrease in HR observed in non-REM sleep was produced by an increase in parasympathetic activity (Gundel, Drescher, Spatenko, & Polyakov, 1999). HR spectral analysis during sleep indicated that HRV modifications during sleep have been related to individual sleep stages and depend on the baseline autonomic HR control level (Zemaityte, Varoneckas, & Sokolov, 1984; Zemaityte, Varoneckas, Plauska, & Kaukenas, 1986). Not only mental effort but also the adaptability of cardiovascular function during the fatigue–restoration cycle can be assessed by HRV (Varoneckas, 2000).

4.1.2.3.3 Blood Pressure

Diastolic and systolic BP have been found to increase due to both mental and physical activity (Boucsein & Backs, 2000). BP is affected by both sympathetic and parasympathetic activity, which results in changes from beat to beat. However, tonic changes in BP are mainly affected by sympathetic activity.

4.1.2.3.4 Respiration

Respiration is not merely a contributing factor to HRV; it can also provide valuable information about operator state. The breathing cycle is characterized by the following parameters: duration of inspiration (Ti), duration of expiration (Te), total cycle time (Ttot), and tidal volume (VT; i.e., the volume that is displaces by one breath). The breathing cycle is centrally controlled by two mechanisms (Wientjes, 1993): a drive mechanism governing the firing rate of the inspiratory neurons, and a timing mechanism switching these neurons on and off. The inspiratory flow rate (VT/Ti) and the timing mechanism index the drive mechanism by the duty cycle time (Ti/Ttot). Mental effort primarily affects the drive mechanism (Wientjes, 1992). Mild stress and mental effort are further associated with an increase in respiratory rate and a decrease in respiratory volume. The respiratory volume is generally increased when the demands are very high. Harding (1987) conducted an extensive study of respiration during high-performance flights. He found different breathing patterns as a function of flight segment. During fighting maneuvers, the respiratory rate was higher and the pilots hyperventilated (more ventilation than required for the situation).

4.1.2.4 Limitations

4.1.2.4.1 What Cardiorespiratory Measures Can Tell Us

Changes in cardiorespiratory parameters often indicate changes in operator functional state. HR increases due to physical and mental activity when the body requires more oxygen (oxygen demand increases during an operational mission). HRV decreases with increasing mental effort and is sensitive to rapid transient shifts in mental work. When the operator is no longer able to cope with high task demands, an increase in HRV may be found, particularly in the 0.10 Hz component. Diastolic and systolic BP values increase due to mental and physical activity: beat-to-beat BP changes are related to the effects of parasympathetic and sympathetic activity, while tonic changes are related to sympathetic control. An increase in HR and a decrease in HRV are due to centrally controlled withdrawal of parasympathetic control during physical work conditions. Increased mental load is accompanied by an increase in sympathetic control (HR and BP increase) followed by suppression of parasympathetic control (HRV decrease).

A decrease in HRV during mental work is accompanied by an increase in respiratory rate (Ti decreases) and a decrease in respiratory volume (VT decreases). Further increased demand for mental effort is accompanied by increases in respiratory volume and inspiratory flow rate (VT/Ti increases). The latter change is followed by an increase in HRV, particularly the LFC. Thus, HRV must be analyzed in parallel with respiratory function. A parallel decrease in both HRV and inspiration flow rate is expected during mild mental load. For increased demands for mental effort, an increase in LFC in parallel with an increased inspiratory flow rate is expected. The difference between those effects might be used for differentiation between efficient mental workload and a performance decrement after working memory has been exceeded. On the other hand, the baseline level of HR frequency and HRV must also be taken into account (Zemaityte, 1989; Kepezenas & Zemaityte, 1983).

Continuous monitoring of BP can be used to analyze baroreflex sensitivity. Baroreflex sensitivity and HRV might be useful for evaluating the impact of mental load because depression of baroreflex sensitivity and reduction of HRV are expected during increased mental effort. The combined use of baroreflex sensitivity, HR, HRV, and inspiratory flow rate might serve as the most efficient indication of the impact of mental effort on cardiorespiratory function.

4.1.2.4.2 What Cardiorespiratory Measures Cannot Tell Us

It is still difficult to use cardiorespiratory measures to gather real-time information because the values are affected by a number of other factors besides operator state.

4.1.2.4.3 Heart Rate Variability

Several real-life studies have demonstrated that HRV alone does not provide adequate information about operator state. It has been found that HRV is higher during task performance than during rests, which is opposite to what might be expected from many laboratory studies that show a relationship between effort investment and HRV (e.g., Aasman et al., 1987). The cardiovascular system is far more complex than the model in Figure 9 suggests. HRV must be interpreted carefully since many other factors besides baroreceptor sensitivity affect HRV. For example, a reduction in the high band is believed to be due to a reduction in parasympathetic activity, which results in reduced baroreceptor sensitivity. Hence, changes in BP are reflected less by changes in HR. Another important factor affecting HRV is respiration. An increase in respiratory frequency and a decrease in respiratory amplitude result in a decrease in HRV. However, the magnitude depends on the respiratory frequency. The largest effect of respiration upon HRV is found when the respiratory frequency is around 0.10 Hz (Angelone & Coulter, 1964). The mid-band region is often assumed to not be affected by respiration in normal situations. However, Veltman and Gaillard (1996, 1998) showed that pilots during simulator flights frequently slow their respiratory frequencies, especially when task demands change. During transitions from high to low task demands and vice-versa, large decreases in respiratory frequency and increases in amplitude are found, resulting in increased HRV, especially in the mid-frequency band. Thus, in order to interpret HRV results adequately, information about respiration is highly desired.

Respiration does not affect the measurement of baroreceptor sensitivity (Veltman & Gaillard, 1996), and therefore this measure provides better information about OFS than HRV alone. However, a disadvantage of this method is that blood pressure must be measured from heart beat to heart beat. This requires sophisticated measurement equipment and real-life measurement becomes more complicated than measuring HR only.

4.1.2.5 General Advantages/Disadvantages of Cardiorespiratory Measures

Most cardiorespiratory measures can be easily measured and provide objective information about operator state while the person is at rest or engaged in a variety of activities.

4.1.2.5.1 Intrusiveness

Because cardiorespiratory measures require sensors to be attached to the body, the intrusiveness is relatively high. Some sensors, such as electrodes for HR and belts for respiration, are not felt by the person. Other sensors, such as cuffs for blood pressure, can directly distract the person and therefore are not recommended for use during task performance.

4.1.2.5.2 Operator Acceptance

There are some cultural differences in operator acceptance with regard to cardiorespiratory measures. Some operators will not agree to be measured because they perceive that deviations in physiological reactions might be harmful to their careers. Therefore, emphasizing that the measures will be used for state assessment only is highly recommended.

4.1.2.5.3 Ease of Use

The ease of use depends upon the desired accuracy of the values to be obtained. For example, average HR can be obtained easily by means of a sports tester. However, to calculate HRV, the accuracy of the HR measurement must be high. Therefore, more sophisticated techniques are required, which reduces the ease of use.

4.1.2.5.4 Artifacts

Much attention should be given to avoid artifacts. HRV is strongly affected by artifacts. ECG artifacts can be due to poor quality of the signal or to incomplete heart beats such as extra systoles. Both types of artifacts can increase HR variability dramatically. Therefore, segments in which artifacts in the ECG signal occur should be corrected using special "correction algorithms" or should not be included in the measurement of HR variability.

4.1.2.6 Apparatus Required

Sophisticated amplifiers and filters are required for most cardiorespiratory measures. However, the quality of the equipment available in recent decades has improved considerably. Many systems allow the user to select several gain and filter settings easily.

4.1.2.7 Personnel Required

Highly qualified personnel are required for cardiorespiratory measurement because most signals are error prone. Furthermore, like most other physiological signals, cardiorespiratory signals can be affected by many other signals. Therefore, interpretation of the signal is often difficult.

4.1.2.8 Analysis Techniques

Electrocardiogram (ECG) recordings can be made from precordial leads, and the occurrence of R-waves in the ECG can be measured electronically using QRS-complex detection by means of a data acquisition personal computer with a sampling rate of at least 100 Hz. Higher sampling frequencies allow more precise measurement of HR and HRV.

HRV can be evaluated by different methods including time-domain (statistical and geometrical), frequency-domain, rhythm pattern analysis, and non-linear methods.

4.1.2.8.1 Time Domain HRV Methods

Heart rate or the inter-beat interval at any point in time (time between successive R-peaks of the ECG) is determined as a starting point. Each QRS complex is detected and the normal-to-normal (R-R) intervals or the instantaneous heart rate is determined. Simple variables can be statistically calculated including:

- Mean R-R interval.
- Standard deviation of the R-R interval (SDRR) – square root of variance (since variance mathematically is equal to the total power of the spectral analysis, SDRR reflects all the cyclic components responsible for variability in the analyzed record).
- HRV triangular index – integral of the density distribution (i.e., the total number of R-R intervals divided by the maximum of the density distribution in order to estimate an overall HRV).
- SDARR – standard deviation of the average R-R interval calculated over short-term periods, usually a stationary process or 5-min period of a long-term record (e.g., 24 hours).
- RMSSD – the square root of the mean of the sum of the squares of differences between adjusted R-R intervals.

The R-R intervals can also be converted into a geometrical pattern such as the sample density distribution of the R-R interval duration, sample density distribution of differences between adjacent R-R intervals, or Lorenz plot(s) of R-R intervals. A simple formula may be used to compute the variability based on the geometric and/or graphic properties of the resulting pattern. Three general approaches are used in geometric methods:

- A basic measurement of the geometric pattern (e.g., the width of the distribution histogram at the specified level) is converted to the measurement of HRV.
- The geometric pattern is interpolated by a mathematically defined shape (e.g., approximation of the distribution histogram by a triangle, or approximation of the differential histogram by an exponential curve), and the parameters of this mathematical shape are used.
- The geometric shape is classified into several pattern-based categories that represent different classes of HRV (e.g., elliptic, linear, and triangular shapes of Lorenz plots). Most geometric methods require the R-R interval sequence to be measured on or converted to a discrete scale that is not too fine or too coarse and that permits the construction of smoothed histograms. Most experience has been obtained with bins approximately 8 ms long (precisely 7.8125 ms=1/128 s), which corresponds to the precision of current commercial equipment.

Two estimates of the overall HRV, SDARR, and RMSSD are recommended because the HRV triangular index permits only justified pre-processing of the ECG signal. All these measurements of short-term variation estimate high frequency variations in HR. The methods expressing overall HRV and its short- and long-term components are not interchangeable and should correspond to the aim of the individual study. Distinction should be made between measures derived from direct measurement of R-R intervals or instantaneous HR, and from the differences between R-R intervals. It is inappropriate to compare time-domain measures, especially those expressing overall HRV, obtained from recordings of different durations.

4.1.2.8.2 Frequency Domain HRV Methods

Power spectral density analysis of HR provides the basic information of how power (variance) is distributed as a function of frequency. Methods for analyzing power spectral density may be classified as non-parametric or parametric. Both methods provide comparable results. Advantages of the non-parametric methods include: (1) simplicity of the algorithm employed (Fast Fourier Transform in most cases), and (2) high processing speed. Parametric methods may be preferred for: (1) smoother spectral components that can be distinguished independently of pre-selected frequency bands, (2) easy post-processing of the spectrum with automatic calculation of individual frequency power components and easy identification of the individual components, and (3) accurate estimation of power spectral density even on a small number (no less 200-300 R-R intervals) of samples across which the signal supposedly maintains stationarity. The main disadvantage of parametric methods is the need to verify the suitability of the chosen model and its complexity (i.e., the order of the model). Power spectral density analysis can be performed either on short-term or long-term R-R interval recordings.

Short-term recordings allow one to analyze frequencies above 0.04 Hz. Measurement of the different frequency components is usually made in absolute units of power (ms^2), but is strongly recommended to be measured in relative values (e.g., to total power) or normalized units (n.u.) that represent relative value of each power component in proportion to the total power minus the VLFC. The representation of LFC and HFC in n.u. emphasizes the controlled and balanced behavior of the two branches of the autonomic nervous system. Nevertheless, relative values or n.u. should always be quoted with total spectral power and absolute values of the individual components in order to describe in total the distribution of power in each spectral component.

Long-term recordings during an entire 24-h period might be used for spectral analysis of R-R intervals. The result then includes an ultra-low frequency component (ULFC) in addition to VLFC, LFC, and HFC. ULFC consists of oscillations in the frequency range ≤0.003 Hz. The slope of the 24-h spectrum can also be assessed on a log-log scale by linear fitting of the spectral values. The problem of "stationarity" is discussed with long-term RR-interval recordings. If the mechanisms responsible for RR-interval modulations of a certain frequency remain unchanged during the entire period of recording, the corresponding frequency component of HRV may be used as a measure of these modulations. If the modulations are not stable, interpretation of the results of the frequency analysis is less well defined. In particular, the physiological mechanisms of the RR-interval modulations responsible for LFC and HFC cannot be considered stationary during the 24-h period. More detailed information about autonomic modulation of R-R intervals is available in shorter recordings demonstrating "stationarity" of the process. It should be remembered that individual oscillatory components provide measurements of the degree of autonomic modulations rather than the level of autonomic tone.

The analyzed ECG signal should satisfy several requirements in order to obtain a reliable HRV calculation and spectral estimation. In order to attribute individual oscillatory components to well-defined physiological mechanisms, such mechanisms modulating HR should not change during the recording. The sampling rate for correct detection of the R wave should be at least 500 Hz. If ectopic beats, arrhythmic events, missing data, or noise effects interrupt "stationarity" of the HR recording, proper interpolation (or linear regression or similar algorithms) using preceding/successive beats of the HR signal or its autoregression function should be used to reduce this error. The relative number and relative duration of R-R intervals that were omitted or interpolated should also be quoted.

4.1.2.8.3 Blood Pressure

BP can be measured using a cuff around the upper arm by means of the Korotkof method. Continuous measurement of finger blood pressure can be obtained by using Finapress or Portapress by means of a cuff around the finger. A continuous method is recommended because it allows the measurement of baroreceptor sensitivity.

4.1.2.8.4 Respiration

The respiratory signal (frequency and amplitude) from the thorax can be recorded using impedance cardiography. Also, for some investigations (e.g., sleep), airflow through the nose and mouth as well as movements of the thorax and abdomen can be measured using appropriate sensors. Because the respiratory frequency is relatively low, the sampling rate of signals can be 10 Hz or higher.

4.1.2.8.5 Oxygen Saturation

Measurement from the finger or auricular lobe is recommended for assessment of OFS when the operator works under specific task demands incorporating the risk of development of hypoxia. For additional details, refer to the section on oxymetry.

4.1.2.9 References

Aasman, J., Mulder, G., & Mulder, L.J.M. (1987). Operator Effort and the Measurement of Heart Rate Variability. *Human Factors*, *29*, 161-170.

Angelone, A., & Coulter, N.A. (1964). Respiratory-Sinus Arrhythmia: A Frequency Dependent Phenomenon. *Journal of Applied Physiology*, *19*, 479-482.

Boucsein, W., & Backs, R.W. (2000). Engineering Psychophysiology as a Discipline: Historical and Theoretical Aspects. In R.W. Backs & W. Boucsein (Eds.), *Engineering Psychophysiology: Issues and Applications* (pp. 3-30). Mahwah, NJ: Lawrence Erlbaum Associates.

Caldwell, J.A., Wilson, G.F., Cetinguc, M., Gaillard, A.W.K., Gundel, A., Lagarde, D., Makeig, S., Myhre, G., & Wright, N.A. (1994). *Psychophysiological Assessment Methods* (AGARD-AR-324). Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development.

Coles, M.G.H., & Sirevaag, E. (1987). Heart Rate and Sinus Arrhythmia. In A. Gale & B. Christie (Eds.), *Psychophysiology and the Electronic Workplace* (pp. 255-274). Chichester, UK: Wiley.

Gundel, A., Drescher, J., Spatenko, Y.A., & Polyakov, V.V. (1999). Heart Period and Heart Period Variability during Sleep on the MIR Space Station. *Journal of Sleep Res*earch, *8*, 37-43.

Harding, R.M. (1987). *Human Respiratory Responses during High Performance Flight* (AGARD-AG-312). Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development.

Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use (1996). Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Circulation*, *93*, 1043-1065.

Hopman, J.C.W., Kollee, L.A.A., Stoelinga, G.B.A., van Geijn, H.P., & van Ravenswaaij-Arts, C.M.A. (1993). Heart Rate Variability. *Annals of Internal Medicine*, *118*, 436-447.

Kalsbeek, J.W.H., & Ettema, J.H. (1963). Continuous Recording of Heart Rate and the Measurement of Mental Load. *Ergonomics*, *6*, 306-307.

Kepezenas, A., & Zemaityte, D. (1983). Relationship between Heart Rate Structure and Physical Fitness in Sportsmen (in Russian). *Fiziologija Cheloveka*, *9*, 729-739.

Kramer, A.F. (1991). Physiological Metrics of Mental Workload: A Review of Recent Progress. In D.L. Damos (Ed.), *Multiple-Task Performance* (pp. 279-328). London: Taylor and Francis.

Malik, M., Bigger, J.T., Camm, A.J., Kleiger, R.E., Malliani, A., Moss, A.J., & Schwartz, P.J. (1996). Guidelines for Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *European Heart Journal*, *17*, 354-381.

Mulder, L.J.M. (1988). *Assessment of Cardiovascular Reactivity by Means of Spectral Analysis*. Thesis. Groningen, The Netherlands: University of Groningen.

Roscoe, A.H. (1992). Assessing Pilot Workload. Why Measure Heart Rate, HRV and Respiration? *Biological Psychology*, *34*, 259-288.

Sirevaag, E.J., Kramer, A.F., Wickens, C.D., Reisweber, M., Strayer, D.L., & Grenell, J.F. (1993). Assessment of Pilot Performance and Mental Workload in Rotary Wing Aircraft. *Ergonomics*, *36*, 1121-1140.

Tattersall, A., & Hockey, G. (1995). Level of Operator Control and Changes in Heart Rate Variability during Simulated Flight Maintenance. *Human Factors*, *37*, 682-698.

Varoneckas, G. (2000). Analysis of Heart Rate Variability during Sleep as a Tool for Assessment of Cardiovascular Adaptability and Fatigue in Sleep-Wake Cycle. In *Individual Differences in the Adaptability to Irregular Rest-Work Rhythms/Status of the Use of Drugs in Sleep-Wakefulness Management* (RTO Meeting Proceedings 31; Venice, Italy 3-4 June, 1999, pp. 22-1 to 22-7).

Veltman, J.A., & Gaillard, A.W.K. (1996). Physiological Indices of Workload in a Simulated Flight Task. *Biological Psychology*, *42*, 323-342.

Veltman, J.A., & Gaillard, A.W.K. (1998). Physiological Workload Reactions to Increasing Levels of Task Difficulty. *Ergonomics*, *5*, 656-669.

Wientjes, C.J.E. (1992). Respiration in Psychophysiology: Methods and Applications. *Biological Psychology*, *34*, 179-204.

Wientjes, C.J.E. (1993). *Psychological Influence upon Breathing: Situational and Dispositional Aspects*. Thesis. Breda, The Netherlands: University of Brabant.

Wierwille, W.W., & Eggemeier, F.T. (1993). Recommendations for Mental Workload Measurement in a Test and Evaluation Environment. *Human Factors*, *35*, 263-281.

Wilson, G.F. (1993). Air-to-Ground Training Missions: A Psychophysiological Workload Analysis. *Ergonomics*, *36*, 1071-1087.

Wilson, G.F., & Eggemeier, F.T. (1991). Psychophysiological Assessment of Workload in Multi-Task Environments. In D.L. Damos (Ed.), *Multiple-Task-Performance* (pp. 329-360). London: Taylor and Francis.

Zemaityte, D., Varoneckas, G., & Sokolov, E. (1984). Heart Rhythm Control during Sleep. *Psychophysiology*, *1*, 279-289.

Zemaityte, D., Varoneckas, G., Plauska, K., & Kaukenas, J. (1986). Components of the Heart Rhythm Power Spectrum in Wakefulness and Individual Sleep Stages. *International Journal of Psychophysiology*, *4*, 129-141.

Zemaityte, D. (1989). Autonomic Control and Development of Complications of Ischemic Heart Disease (in Russian). *Fiziologija Cheloveka*, *15*, 3-13.

4.1.3 Core Temperature

4.1.3.1 Description of Core Temperature

The measurement of body core temperature is one of the most frequently applied diagnostic techniques in medical science (Bartels & Bartels, 1991). This is due to the fact that temperature is a very important indicator for health problems. Core temperature is also the "gold standard" measurement for monitoring the circadian cycle. It has also been used to monitor the effects of physically demanding jobs and the effects of wearing special equipment such as chemical defense gear. Core temperature is the method of choice for these applications because alternative measurements on or near the body's surface may vary greatly since they are influenced by a number of environmental factors.

4.1.3.2 Background

Core temperature in a resting-state is normally in the range of 36.5°C to 37°C. An increase in core temperature in conjunction with a decrease in skin temperature due to sweating is induced by physical work. The core temperature is kept constant during work over a longer period of time as long as the fluid loss is compensated by drinking. Dehydration leads to an increase in core temperature and therefore reduces physical capability.

The term core temperature implies that there is a single temperature value within the core of the human body. However, there are differences within the core of 0.2°C to 1.2°C. The highest temperatures are usually measured in the area surrounding the rectum. Therefore, it is not possible to express core temperature with a single value.

Core temperature follows a regular, day-periodic variation with an amplitude of about 1°C (Figure 10). Humans exhibit a temperature minimum in the early morning at 0400 and a maximum, usually consisting of two peaks around 1800 and 2000 in the evening. By means of environmental pacemakers this periodic variation is usually synchronous with the 24-hour diurnal rhythm. In the absence of these synchronizers the cycle duration of the endogenous, circadian rhythm is about 24 to 25 hours. Strong phase shifts of the external diurnal rhythm (e.g., due to trans-meridian flights) cause temporary desynchronization of the internal circadian rhythm.

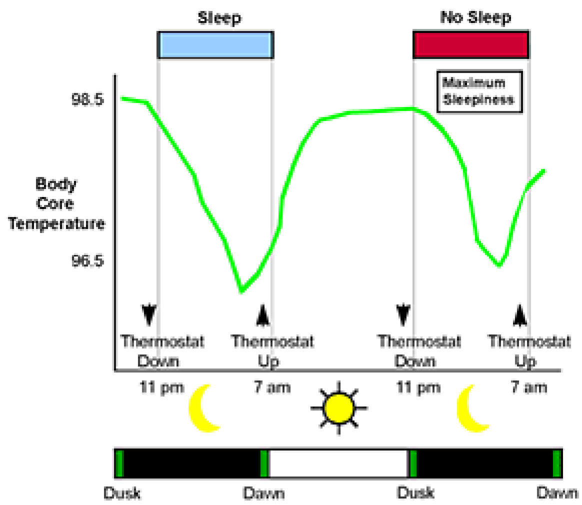

Figure 10: Circadian Rhythm of Core Temperature (Circadian Technologies, Inc., 2002).

Another periodic variation applies to the core temperature of women with an intact ovulation cycle. Shortly after ovulation (within 2 days), there is an increase in basal temperature (which is measured in the morning before getting up and without a physical load) of approximately 0.5°C, which is maintained until menstruation (i.e., for about 14 days). During pregnancy the temperature remains at the increased level.

Variations of environmental temperature below the upper critical temperature of about 30°C (resting state, unclothed) do not cause variations in core temperature but lead to changes in the area in which core temperature can be accurately measured (Schmidt & Thews, 1995). As shown in Figure 11, the area represented as core temperature decreases in cold and increases in warm environments. Environmental temperatures above the upper critical temperature (heat stress) cause an increase of core temperature and initiates heat dissipation by means of sweating. Core temperatures above 39.5 to 40°C pose high strain upon the metabolism and the circulation. Short-term increases up to 42°C are usually not life threatening. The regulation range of temperature has an upper boundary that is reached if the heat dissipation is not sufficient for thermal balance. This value strongly depends upon relative humidity such that the value is higher with low humidity (e.g., 50°C at 30% relative humidity) than with high humidity. A slight heat stress may cause heat inanition (circulation feebleness) or heat collapse (blood pressure feebleness). If the thermal regulation system is overstrained, this leads to hyperthermia with perilous heat prostration causing cerebral symptoms such as unconsciousness and seizures.

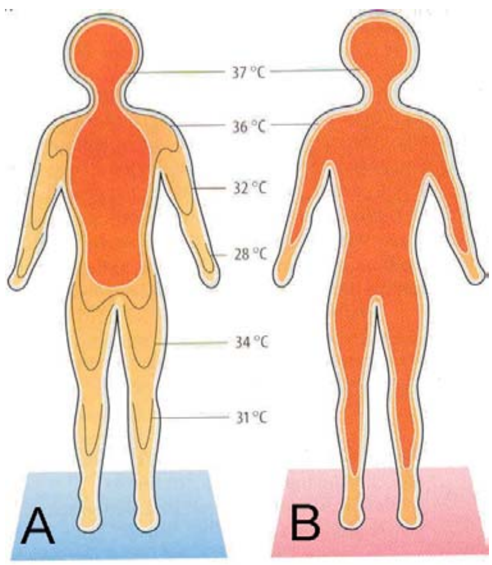

Figure 11: Temperature Distribution of the Human Body in Cold (A; 20°C) and Warm (B; 35°C) Environments (Schmidt & Thews, 1995).

4.1.3.3 State of the Art

The most accurate measure of core temperature is said to be from the vicinity of the pulmonary artery. This requires the use of a thermistor catheter, which is, of course, an invasive procedure. Because pulmonary artery measurement is not practical, other procedures are typically used. These include:

- *Rectal Temperature*: Temperature measured in the rectum can vary by 1°C depending upon the depth that sensor is inserted. Therefore, it is important for intra- and inter-individual comparisons that the sensor is placed at a standard depth.
- *Sublingual Temperature*: The temperature is measured in the mouth cavity with the sensor placed under the tongue. This produces temperatures that are 0.2°C to 0.5°C below the rectal temperature and that are affected by the temperature of inhaled air, drinks, and food.

- *Esophagus* (O*esophagus*) *Temperature*: This measure of temperature is from a sensor placed in the esophagus above the crossover from the gullet to the stomach.
- *Axillary Temperature*: Temperature is measured by placing the sensor in the armpit or groin. The upper arm is pressed against the chest wall or the legs are kept together to achieve stable readings. Time delays of about 10 minutes must be tolerated before the temperature measured reflects the core temperature.
- *Epitympanel Temperature*: Temperature is measured in the outer auditory passage close to the eardrum. Both thermistor and infrared sensors can be used.
- *Telemetry*: A small pill-shaped biotelemeter is swallowed. The device monitors temperature from the digestive tract and telemeters the signal to a receiving device outside the body. The transmitter can either be swallowed so that it moves along the digestive tract within 26 to 80 hours or it can be placed at a fixed position inside the body by means of endoscopy.
- *Topical Strips*: Flexible plastic strips with chemical dots that change color can also be used. However, accuracy issues must be addressed because these sensors are highly susceptible to ambient temperature conditions as well as sweating and vasodilatation. Their accuracy must also be checked.

With regard to the measurement techniques mentioned above, only the measurement of epitympanel temperature can be considered as a technique with adequate accuracy that is agreeable to the subject. Some of the other techniques show either inaccuracy due to poor sensor positioning (e.g., axillary and sublingual temperature), or are uncomfortable (e.g., rectal and esophageal temperature). Portable devices are available for all of the techniques mentioned.

Biotelemetric techniques provide continuous, accurate, and convenient measurement. This seems to be the best practical option, especially if the sensor is positioned within the gastrointestinal tract, because this requires no medical invasion. Because this technique was recently developed in connection with the NASA space program, there is only a single supplier at this time. With this technique, the sensor (thermometer) is packaged within a capsule coated with silicone, which is 22 mm in length and 9 mm in diameter (see Figure 12). The capsule contains a thermo-sensitive crystal, accurate within 0.1°C, whose oscillation frequency depends upon the environmental temperature (i.e., the core temperature). The crystal's oscillation creates a magnetic flux that is picked up, magnified, and transmitted by electronic components inside the capsule. Power is supplied by a small battery. The capsule is swallowed by the subject and leaves the gastrointestinal tract within 26 to 80 hours depending upon individual characteristics. The transmitted signals are received by a small monitoring device that can be placed up to 76 cm from the sensor. The receiver is compatible with personal computers so that the temperatures can be continuously recorded and analyzed.

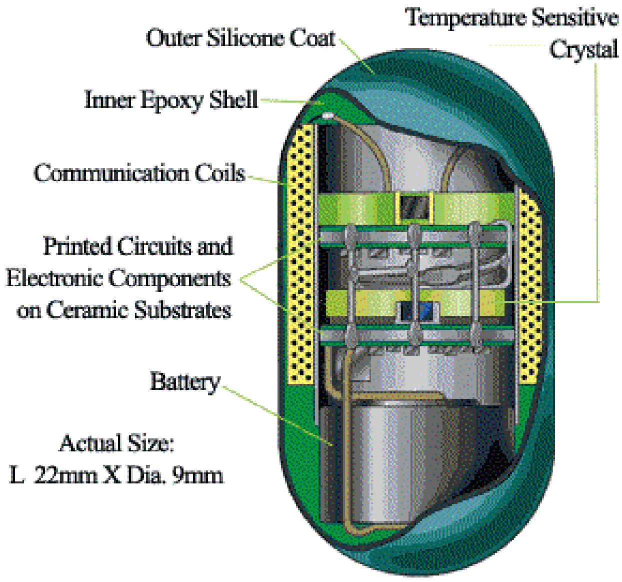

Figure 12: CorTemp™ Thermometer Pill (left) and Ambulatory Recorder (right) (HTI Technologies, Inc., 2002).

4.1.3.4 Limitations

4.1.3.4.1 What Core Temperature Can Tell Us

The measurement of core temperature provides useful information about the physiological state of the human body, especially in connection with the effects of shift work and jet lag on the circadian rhythm, which can harm performance. Core temperature can also be used to examine the effects of heat stress and physical work. Core body temperature is the gold standard for monitoring circadian rhythms.

4.1.3.4.2 What Core Temperature Cannot Tell Us

Evaluation of data should take into account the type of sensor, the sensor's location in terms of placement within the body, and the time of day with respect to circadian rhythm effects. Ambient temperature and temperature gradients should also be addressed when interpreting the results. A single measurement is not sufficient for the purpose of circadian rhythm evaluation. Several measurements should be taken within an hour. Since several factors can influence absolute measures of core temperature, baseline values can be recorded under standard conditions and used to derive relative temperature measurements.

4.1.3.5 General Advantages/Disadvantages of Core Temperature

There is no measurement technique for core temperature that is completely non-invasive to the subject. Either a sensor must be positioned at sensitive spots of the human body or the sensor must be swallowed, which might be uncomfortable to some subjects.

4.1.3.6 Apparatus Required

The apparatus that is required for the measurement depends upon the technique that is applied.

4.1.3.7 Personnel Required

For the measurement at least one person is needed who performs the positioning of the sensor and maintains the apparatus.

4.1.3.8 Analysis Techniques

No special statistical techniques are needed to analyze temperature data. For determining circadian rhythms, software is available to smooth the data and to determine the phase of the rhythm. This analysis is useful to determine the phase of the body in a new time zone when time zones are crossed or to study the effects of shift work.

4.1.3.9 References

Bartels, H., & Bartels, R. (1991). *Physiologie: Lehrbuch und atlas*, *4*. Auflage. München: Urban und Schwarzenberg.

Circadian Technologies, Inc. (2002). Circadian Rhythms and Your Biological Clock – Circadian Temperature Rhythm. Download from http://www.circadian.com/learning_center/circa-dian_temp.htm at 28.01.2002.

HTI Technologies, Inc. (2002). CorTemp™ Wireless Monitoring System. Download from http://www.htitech.com/Personal-Safety.htm at 28.01.2002.

Schmidt, R.F., & Thews, G. (1995). *Physiologie des menschen*, *26*. Auflage. Berlin: Springer.

4.1.4 Electroencephalography (EEG)

4.1.4.1 Description of EEG

The electrical activity of the human brain was first recorded as the electroencephalogram (EEG) by Berger (1930). He reported changes in the pattern of the EEG between periods of mental engagement and resting. The EEG was first monitored during sleep in 1935 (Loomis, Harvey, & Hobart, 1935). The EEG has become the gold standard for classifying and studying the stages of sleep (Rechtschaffen & Kales, 1968). The brain receives sensory information, is responsible for processing this information, then makes decisions and initiates action. Because of this, the EEG is used as a measure of brain engagement in cognitive tasks. The EEG spectra can be analyzed to determine the levels of involvement present during different types of cognitive activity (Davidson, Jackson, & Larson, 2000; Wilson & Eggemeier, 1991).

4.1.4.2 Background

There is a large body of literature showing the relationship between EEG and cognitive activity (Davidson, Jackson, & Larson, 2000; Wilson & Eggemeier, 1991). EEG is routinely used in clinical medicine to diagnose numerous disease states. The use of EEG in human factors and ergonomics settings is not as widespread (Caldwell et al., 1994). One reason is because of the difficulty of measuring EEG signals in non-laboratory conditions. However, the currently available technology has overcome this obstacle. Because the EEG is a complex waveform, signal-processing techniques are required to process the data. Typically, spectral analysis is used to analyze the EEG data and the data are partitioned into five bands. The bands are delta, theta, alpha, beta, and gamma, from slowest to fastest. The power in each band is computed and used to compare the conditions being studied. Event-related brain potentials (ERP) have found widespread application in many research laboratories and clinics. They monitor small changes in the EEG while the brain is processing information related to specific external or internal events. Because of the small amplitude of the ERPs relative to the much larger background EEG, response averaging is required to extract the ERPs. This requires presentation of the eliciting conditions or stimuli a number of times in order to extract the ERPs. This requirement for repeating the stimuli has hampered the use of ERPs in applied settings because it requires more time than is usually available.

4.1.4.3 State of the Art

In human factors applications, data are collected while operators perform their jobs or under simulated work conditions. An example of this approach is shown in a simulated air traffic control task. The effects of three levels of mental workload during a simulated air traffic control experiment were determined by recording several psychophysiological measures including EEG from 19 scalp sites (Brookings, Wilson, & Swain, 1996). Only the EEG measures were able to discriminate among the different manipulations of mental workload. In a further analysis of the EEG from this study, Russell and Wilson (1998) used a neural network classifier to discriminate between the levels of mental workload. They reported a mean correct classification accuracy discriminating among the workload levels of 84%. When the discrimination was between only the overload condition and the other conditions, a 92% classification accuracy was achieved. This suggests that the level and nature of mental workload can be accurately determined using EEG data and advanced classification techniques.

EEG was recorded in a low-fidelity simulation task by Sterman, Mann, Kaiser, and Suyenobu (1994) to determine changes in topography of the EEG activity when subjects performed simulated aircraft landings. They found reduced alpha band activity over central and parietal scalp sites during the landings. They interpreted the reduction at the parietal scalp sites as being related to the cognitive processing required by the landing task.

EEG has also been used to study graded levels of mental workload during actual flight. Wilson (2002) recorded 29 EEG channels during a 90-minute flight. The results from this study showed that EEG alpha band activity decreased over the right posterior sites during the more demanding instrument flight rule (IFR) flight segments and during visual flight rule landings, missed approaches, and climb outs. The delta band EEG activity at central and parietal scalp sites showed increased activity during touch-and-go, landings, take offs, and the IFR segments.

Most continuous or sustained operations lead to complaints of fatigue, sleepiness, sleep disturbances, and decreased performance. These conditions can lead to an increased performance error rate and mission failure. Results of sleep deprivation studies in healthy subjects support the relation between sleepiness and memory deficiency (Dinges & Kribbs, 1991). Even modest reductions in sleep time are associated with cognitive performance impairment (Blagrove, Alexander, & Horne, 1994).

The factors that regulate fatigue, alertness, performance, and thus operator functional state are the circadian and sleep-wake systems (Folkard & Akerstedt, 1989). Regarding sleep quantity, partial or total sleep deprivation is followed by increased daytime sleepiness the following day (Bonnet, 1985, 1986; Carskadon & Dement, 1982). Even sleep fragmentation without awakening (by inducing repeated microarousals) is sufficient to increase sleepiness (Stepanski, Lampere, Roehrs, Zorick, & Roth, 1987). Therefore, modest nightly sleep deprivation accumulates over nights to progressively increase daytime sleepiness and performance lapses (Carskadon & Dement, 1981). Conversely, increasing sleep time beyond the usual 7-8 h per night decreases sleepiness (Roehrs & Carskadon, 1998). Night train driving has been associated with increased subjective sleepiness, and increased levels of slow eye movements, alpha, theta, and delta band EEG activity. Four of the night drivers admitted dozing off and two drivers missed signals during times of high-amplitude alpha bursts (Torsvall & Akerstedt, 1987).

The effects of space flight on sleep were investigated using EEG during a 30-day mission (Gundel, Polyakov, & Zulley, 1997). Normal sleep patterns, as determined by ground data, were disrupted by space flight and the circadian phase was delayed by 2 hours while the latency to rapid eye movement sleep was shorter. Drugs to induce and avoid sleep act on the central nervous system and produce changes in the EEG. The pattern of changes in the EEG can be used to determine the effects of these drugs (Caldwell et al., 1994).

4.1.4.4 Measurement of Sleep

EEG is the "gold standard" for quantifying sleepiness, and for measuring sleep quality and quantity. Sleep can be objectively measured using standard polysomnographic recordings that usually include four EEG channels, electrooculography of each eye (oblique and horizontal derivations), and chin electromyography. EEG signals are provided with electrodes attached to the scalp with collodion, or held in place with an electrode cap. Four scalp electrode sites are typically used and are positioned over the following 10-20 sites: C3, Cz, O1, and O2, related to a reference (A1) on the left ear or mastoid process. This layout provides a good evaluation of sleep in healthy subjects, even during field studies (Beaumont et al., 2000). After amplification and filtering (Epstein, 1993), all signals can be either directly read on a computer or stored using a portable recorder to be analyzed later according to the standard criteria developed by Rechtschaffen and Kales (1968).

4.1.4.5 Measurement of Sleepiness

4.1.4.5.1 Continuous EEG Recording

Sleepiness is measured by scoring micro-sleep episodes or by using a computerized quantitative analysis of EEG waveform features. Micro-sleep episodes are indicated by increased amounts of alpha and theta activity in behaviorally awake humans deprived of or restricted from sleep (Akerstedt, Torsvall, & Gillberg, 1982). Precise assessment of the degree of sleepiness is provided by accumulating all micro-sleep episodes over a work period. Delta activity has been reported to increase in response to experimental sleep deprivation while alpha activity decreased (Borbély, Baumann, Brandeis, Strauch, & Lehman, 1981). Whereas micro-sleep episodes can be directly counted, quantitative sleep EEG analysis requires specific software or a trained analyst.

4.1.4.5.2 Intermittent EEG Recording

The Multiple Sleep Latency Test (MSLT) represents the standard physiological tool for quantifying sleepiness (Carskadon, Dement, Mitler, & Roth, 1986). The MSLT is determined at 2-h intervals throughout the day. The primary measure of the MSLT is the latency to fall asleep while lying with eyes shut in a quiet, dark room. Sleep latencies in healthy normal individuals range from 10 to 20 min. Sleepiness is defined as a mean sleep latency of less than 5 or 6 min (Carskadon & Dement, 1981).

The Maintenance Wakefulness Test (MWT) can also be used to assess sleepiness. The MWT tests the ability to remain *awake* by measuring the latency to sleep onset. The test employs 4 to 6 sessions lasting 20, 30 or 40 min, scheduled at 2-h intervals, beginning 2 hours after awakening from the previous night's sleep (Mitler, Gujavarty, & Browman, 1982). The subjects lie in bed or sit in a chair in a darkened room.

4.1.4.6 Limitations

4.1.4.6.1 What EEG Can Tell Us

EEG measures of cognitive task activity can be very useful for determining the functional state of operators. Characteristic EEG patterns comprise a very wide range of brain activity from deep sleep at one end of the continuum to high mental workload levels at the other end. Various types of cognitive activity can be determined using EEG spectra and additionally by examining the topography. Because the brain is the seat of thinking, the EEG provides direct measures of brain activity during cognitive activity.

Sleep quality and quantity can be measured with continuous EEG recording using an electrode cap, and sleepiness can be evaluated with the MSLT (Beaumont et al., 2000; Lagarde et al., 2000). The advantages of the MSLT are its direct, objective, quantitative approach, the availability of normative values and test standardization, and its reliability. In contrast to tests of performance, motivation does not seem to reduce the impact of sleep loss as measured by the MSLT (Hartse, Roth, & Zorick, 1982). The MWT can also be used to measure the ability to resist sleep or the ability to not be overwhelmed by sleepiness, thus extending the sensitivity range of the MSLT.

In operational conditions, sleepiness can only be assessed using portable techniques. Continuous EEG using a portable recording unit provided with long-life batteries permits the scoring of micro-sleep in subjects while they are performing tasks (Lagarde et al., 2000).

4.1.4.6.2 What EEG Cannot Tell Us

Currently, technology and theory do not permit a fine-grained analysis of specific cognitive mechanisms using the EEG. EEG can be used to determine whether or not an operator is experiencing cognitive overload but probably not what is causing the overload. That is, it is possible to correctly classify the overload state, but current analysis techniques do not provide an assessment of the specific cognitive conditions causing the overload condition.

Although fatigue is well correlated with sleepiness and sleep, EEG appears to be an indirect tool for assessing fatigue. Except for continuous EEG, EEG techniques must be used in standardized conditions (Roehrs & Carskadon, 1998) in order to eliminate the influence of environmental stressors and potential risk factors. Characteristics of the individual such as personality are not taken into account.

4.1.4.7 General Advantages/Disadvantages of EEG

The EEG technique represents the gold standard objective tool for measuring sleep and sleepiness. However, fitting the subject with electrodes requires at least half an hour and one technician. Data processing and interpretation of data requires additional time and expertise. Moreover, the subject's tolerance for the recording device decreases with time. Monitoring conditions are also restrictive. For the MSLT, subjects are not permitted to remain in bed between nap test sessions, which can disturb the recovery sleep scheduled at a given period of time. Moreover, subjects should not engage in vigorous pre-test activity because it will alter the test outcome (Bonnet & Arand, 1998). The room must be dark and quiet during testing; polysomnographic parameters (EEG, EOG, and chin muscle EMG) needed to detect sleep onset and score sleep stages must be recorded during nap opportunities.

A major criticism of MWT relates to the wide variety of protocols used. The length of the test session has not been well standardized (20, 30, or 40 min). Also, normative data exist only for a 40-min test in healthy subjects; for other session lengths, projected norms are used, but they have not been validated (Doghramji et al., 1997). MWT appears to be less sensitive than MSLT in determining sleepiness. Studies comparing sleep latencies measured using MWT and MSLT have shown that one takes longer to fall asleep when instructed to remain awake (MWT) than when told not to resist sleep (MSLT; Sangal, Thomas, & Mitler, 1992). Other issues include the effects of age, sleep deprivation, time of testing, and drugs on MWT profiles.

EEG spectral analysis is not often used for assessing sleepiness for several reasons, including the need for trained technicians. The lack of standardization and the absence of normative values also limit its usefulness. The high degree of between-subject variation makes it difficult to compare results between individuals. Finally, EEG measures may not be as sensitive to episodes of sleepiness and as predictive of performance lapses as continuous video monitoring (Dinges & Mallis, 1998).

The EEG is sensitive to numerous artifacts that are abundantly present in the work environment. These artifacts include eye movements, muscle activity, body and limb movement, and electrical interference. However, there are numerous signal-processing techniques available to detect and remove the artifacts. This is especially beneficial in human factors work where important events often cannot be repeated. If the data from a critical event are contaminated with artifacts, they need not be discarded if the artifact can be removed.

4.1.4.8 Apparatus Required

All the EEG techniques require an EEG recording unit, electrodes, abrasive paste, and adhesive paste. Additionally, a PC computer fitted with appropriate software is necessary for calibration, programming, downloading, and analyzing the data. The EEG measures the difference in electrical potential between pairs of electrodes placed on the scalp. These signals are amplified and then filtered to produce an analog or digital recording (Epstein, 1993). The international 10-20 system of EEG electrode placement or a derivative with additional electrodes is customarily used. The 10-20 refers to 10% and 20% of the distances between standard cranial landmarks (Keenan, 1994). Each recording channel is derived from the signals from a pair of electrodes or a number of electrodes electrically referenced to an electrically neutral site and combined to form a montage. Any number of channels are typically monitored from one to more than 256 in research laboratories. Typically, field studies record from a small number of electrode sites. When studying sleep, electrooculography and electromyography channels are also recorded and, taken together, constitute polysomnography. At least two EEG channels are usually used for polysomnography recordings.

4.1.4.9 Personnel Required

The EEG technique requires the presence of skilled technicians or physicians. The personnel need to test the recorders, apply electrodes, monitor the data recording, and analyze the data.

4.1.4.10 Analysis Techniques

EEG is typically analyzed using readily available software. The analysis usually includes detecting and removing artifacts, performing spectral analysis, and applying appropriate statistical analyses. Specific software compatible with the recorders is needed to analyze sleep and sleepiness. Nevertheless, due mainly to a relative imprecision in standard analysis criteria, a skilful technician or physician must review the analysis of the data, which can sometimes result in a modification of some sleep parameters.

4.1.4.11 References

Akerstedt, T., Torsvall, L., & Gillberg, M. (1982). Sleepiness and Shiftwork: Field Studies. *Sleep*, *5*, S95-S106.

Beaumont, M., Denis, J.B., Piérard, C., Batéjat, D., Van Beers, P., Chauffard, F., French, J., & Lagarde, D. (2000). Slow Release Caffeine Alleviates Jet Lag-Related Sleep Perturbations and Daytime Sleepiness after a 7-time Zone Eastbound Flight. *Aviation, Space, and Environmental Medicine*, *71*, 298.

Berger, H. (1930). Uber das Elektrenkephalogramm des Menschen. *Journal of Psychology and Neurology*, *40*, 160-179.

Blagrove, M., Alexander, C., & Horne, J.A. (1994). The Effects of Chronic Sleep Reduction on the Performance of Cognitive Tasks Sensitive to Sleep Deprivation. *Applied Cognitive Psychology*, *9*, 21-40.

Bonnet, M.H. (1985). The Effect of Sleep Disruption on Performance, Sleep and Mood. *Sleep*, *8*, 11-19.

Bonnet, M.H. (1986). Performance and Sleepiness as a Function of the Frequency and Placement of Sleep Disruption. *Psychophysiology*, *23*, 263-271.

Bonnet, M.H., & Arand, D.L. (1998). Sleepiness as Measured by Modified Multiple Sleep Latency Test Testing varies as a Function of Preceding Activity. *Sleep*, *21*, 477-483.

Borbély, A.A., Baumann, F., Brandeis, D., Strauch, I., & Lehman, D. (1981). Sleep Deprivation: Effect on Sleep Stages and EEG Power Density in Man. *Electroencephalography and Clinical Neurophysiology*, *51*, 483-493.

Brookings, J.B., Wilson, G.F., & Swain, C.R. (1996). Psychophysiological Responses to Changes in Workload during Simulated Air Traffic Control. *Biological Psychology*, *42*, 361-378.

Caldwell, J.A., Wilson, G.F., Centigue, M., Gaillard, A.W.K., Gundel, A., Lagarde, D., Makeig, S., Myhre, G., & Wright, N.A. (1994). *Psychophysiological Assessment Methods* (AGARD Advisory Report AGARD-AR-324). Paris, France: NATO AGARD.

Carskadon, M.A., & Dement, W.C. (1981). Cumulative Effects of Sleep Restriction on Daytime Sleepiness. *Psychophysiology*, *18*, 107-113.

Carskadon, M.A., & Dement, W.C. (1982). Nocturnal Determinants of Daytime Sleepiness. *Sleep*, *5*, S73-S81.

Carskadon, M.A., Dement, W.C., Mitler, M.M., & Roth, T. (1986). Guidelines for the Multiple Sleep Latency Test (MSLT): A Standard Measure of Sleepiness. *Sleep*, *9*, 519-524.

Davidson, R.J., Jackson, D.C., & Larson, C.L. (2000). Human Electroencephalography. In J.T. Cacciopo, L.G. Tassinary, & G.G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 27-52). Cambridge: Cambridge University Press.

Dinges, D.F., & Kribbs, N.B. (1991). Performing While Sleepy: Effects of Experimentally-Induced Sleepiness. In T. Monk (Ed.), *Sleep, Sleepiness and Performance* (pp. 97-128). New York: John Wiley and Sons.

Dinges, D.F., & Mallis, M.M. (1998, February 9-13). Managing Fatigue by Drowsiness Detection: Can Technological Promises be Realized? Third International Conference on Fatigue in Transportation: Coping with the 24-h Society, Fremantle, Western Australia.

Doghramji, K., Mitler, M.M., Sangal, R.B., Shapiro, C., Taylor, S., Walsleben, J., Belisle, C., Erman, M. K., Hayduk, R., Hosn, R., O'Malley, E.B., Sangal, J.M., Schutte, S.L., & Youakim, J.M. (1997). A Normative Study of the Maintenance of Wakefulness Test (MWT). *Electroencephalography and Clinical Neurophysiology*, *103*, 554-562.

Epstein, C.M. (1993). Technical Aspects of EEG: An Overview. In E. Wyllie (Ed.), *The Treatment of Epilepsy: Principles and Practices* (pp. 202-210). Philadelphia, PA: Lea and Febiger.

Folkard, S., & Akerstedt, T. (1989). Towards the Prediction of Alertness on Abnormal Sleep/Wake Schedules. In A. Coblentz (Ed.), *Vigilance and Performance in Automated Systems* (pp. 287-296). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Gundel, A., Polyakov, V.V., & Zulley, J. (1997). The Alteration of Human Sleep and Circadian Rhythms during Spaceflight. *Journal of Sleep Research*, *6*, 1-8.

Hartse, K.M., Roth, T., & Zorick, F.J. (1982). Daytime Sleepiness and Daytime Wakefulness: The Effect of Instruction. *Sleep*, *5*, S107-S118.

Keenan, S.A. (1994). Polysomnographic Technique: An Overview. In S. Chokroverty (Ed.), *Sleep Disorder Medicine* (pp. 84). Boston, MA: Butterworth-Heinemann.

Lagarde, D., Batéjat, D., Sicard, B., Trocherie, S., Chassard, D., Enslen, M., & Chauffard, F. (2000). Slow Release Caffeine: A New Response to the Effects of a Limited Sleep Deprivation. *Sleep*, *23*, 651-661.

Loomis, A.L., Harvey, E.N., & Hobart, G. (1935). Potential Rhythms of the Cerebral Cortex during Sleep. *Science*, *81*, 597-598.

Mitler, M.M., Gujavarty, K.S., & Browman, C.P. (1982). Maintenance of Wakefulness Test: A Polysomnographic Technique for Evaluating Treatment Efficacy in Patients with Excessive Somnolence. *Clinical Neurophysiology*, *53*, 658-661.

Rechtschaffen, A., & Kales, A. (Eds.). (1968). A Manual of Standardized Terminology. Techniques and Scoring System for Sleep Stages of Human Subjects. Washington, DC: U.S. Government Printing Office.

Roehrs, T.A., & Carskadon, M.A. (1998). Standardization of Methods: Essential to Sleep Science. *Sleep*, *21*, 445.

Russell, C.A., & Wilson, G.F. (1998). Air Traffic Controller Functional State Classification using Neural Networks. *Proceedings of Artificial Neural Networks in Engineering (ANNIE'98) Conference*, *8*, 649-654.

Sangal, R.B., Thomas, L., & Mitler, M.M. (1992). Disorders of Excessive Sleepiness: Treatment Improves Ability to Stay Awake but does not Reduce Sleepiness. *Chest*, *102*, 699-703.

Stepanski, E., Lampere, J., Roehrs, T., Zorick, F., & Roth, T. (1987). Experimental Sleep Fragmentation in Normal Subjects. *International Journal of Neurosciences*, *33*, 207-294.

Sterman, M.B., Mann, C.A., Kaiser, D.A., & Suyenobu, B.Y. (1994). Multiband Topographic EEG Analysis of a Simulated Visuomotor Aviation Task. *International Journal of Psychophysiology*, *16*, 49-56.

Torsvall, L., & Akerstedt, T. (1987). Sleepiness on the Job: Continuously Measured EEG Changes in Train Drivers. *Electroencephalography and Clinical Neurophysiology*, *66*, 502-511.

Wilson, G.F. (2002). An Analysis of Mental Workload in Pilots during Flight using Multiple Psychophysiological Measures. *International Journal of Aviation Psychology*, *12*, 3-18.

Wilson, G.F., & Eggemeier, F.T. (1991). Psychophysiological Assessment of Workload in Multitask Environments. In D.L. Damos (Ed.), *Multiple Task Performance* (pp. 329-360). London: Taylor and Francis.

4.1.5 Electrodermal Activity

4.1.5.1 Description of EDA

Electrodermal activity (EDA) or changes in the electrical characteristics of skin is one of the most widely known and used electrophysiological indices. The electrodermal response (EDR) arises as a response to discrete stimuli. The time course of EDR can take up to tens of seconds. The latency of the onset of the EDR is typically one to three seconds following the eliciting stimulus. Typically, the conductance of the skin is measured. There are two components, the skin conductance response (SCR), which is the phasic response to a stimulus and the skin conductance level (SCL), which is a measure of the tonic level of skin conductance and has a time course of up to several minutes (Andreassi, 2000; Dawson, Schell, & Filion, 2000; Stern, Ray & Quigley, 2001).

4.1.5.2 Background

EDA is associated with eccrine sweat gland activity and is most readily recorded from the palms of the hands and the soles of the feet. Eccrine sweat gland activity is driven by both the need to cool the body and in response to emotional and cognitive situations. The heat regulative function is realized in general by a continuous excretion that leads to tonic changes of the skin resistance.

The EDA can be assessed as a change in potential, a change in resistance, or a change in conductance. Typically, the skin resistance is measured and converted to conductance units. Resistance is measured by passing a constant current between two electrodes attached to the hand or the foot. Changes in resistance are determined over time. Another method measures the difference in the electrical potential between the two electrodes (Dawson et al., 2000).

Three types of responses can be measured. The first is the tonic level, or skin conductance level (SCL), which is seen as very slow changes that can occur over minutes. This measure provides an indication of the background level of the operator. The phasic response to discrete stimuli, the skin conductance response (SCR), has a latency of one to three seconds following the onset of the eliciting stimulus. Spontaneous changes in skin conductance also occur and can be used to characterize operators as labile or non-labile.

EDA has been used for decades to monitor the effects of arousal, anxiety and attention. It has been used to measure the significance of stimuli in many different contexts (Dawson et al., 2000). Although most of the research using EDA has been in laboratory and clinical settings, there have been applications in applied settings (Backs & Boucsein, 2000).

4.1.5.3 State of the Art

EDR appears to be under the influence of emotions and has a low correlation with heat regulation. Aldersons (1985) reported that there are six phases of the EDR. The first phase is a thermal effect (TE), the second phase relates to physical work (PW) and the third phase relates to mental work (MW). He suggested that the relationship among the phases was “TE -> PW -> MW”. The fastest changes appear during mental work. At the same time, emotional effort is not accompanied by high values of EDA. Both EEG and EDA demonstrate the same changes in the same direction over a quite narrow range. The period of relative stability is accompanied by 0.2-0.5 Hz rhythms and coincides with very slow potential rhythms of the brain (Alajalova, Leonova, & Rusalov, 1975).

To eliminate the phasic structural ambiguity of the EDA, Karpenko et al. (1984) proposed using its integral. The dynamics of the integral correspond to the oscillations in performance in humans during different functional states (Burov, 1986).

EDA has been shown to demonstrate sensitivity to both emotional and cognitive events (Backs & Boucsein, 2000). The vast majority of research using EDA has been with laboratory tasks. A much smaller number of studies have used EDA in the work environment. Some of these results are shown in Table 11 where the direction of the electrodermal response change is shown.

Table 11: Changes in SCR Parameters Associated with various Environmental Events

Measure	Sensitivity	Study
SCR amplitude ↑	Short-term workload during approach in simulated flight	Lindholm and Cheatham (1983)
NS.SCR frequency ↑	Emotional load by prolonged SRTs and time pressure	Schaefer et al. (1986) Kuhmann et al. (1987)
NS.SCR frequency ↑	Perceived difficulty of road curvature during car driving	Richter et al. (1998)
SCR amplitude ↑	Increased emotional load during car driving	Heino et al. (1990)
SCR amplitude ↑	Probability of an aversive event	Backs and Grings (1985) Lovibond (1992)
SCR amplitude ↑	User-hostile task structure in HCI	Muter et al. (1993)
Integrated SCR amplitude ↑	Developing emotional strain	Kuhmann (1989)
NS.SRR frequency ↓	Decreasing emotional load under long system response times in HCI when time pressure is absent	Kuhmann et al. (1990)
NS.SCR frequency ↓	After prolonged simulated night shift work and prolonged work under noise	Boucsein and Ottmann (1996)
Integrated SCR amplitude ↓	Three hours of perceptual task performance when time pressure is absent	Burov (1986)

Note: "↑" indicates that the values of the parameter in question increase with increasing strain, "↓" that they decrease.

HCI = human-computer interaction SRT = system response time

SCR = skin conductance response NS.SCR = non-specific skin conductance response

SRR = skin resistance response NS.SRR = non-specific skin resistance response

(Adapted from Backs & Boucsein, 2000)

Boucsein (2000) summarized his work with office workers performing complex computer tasks. Using EDA and other psychophysiological measures, he concluded that in this situation rest breaks should be given according to the needs of the operators and the task demands rather than on the basis of a rigid rest break schedule. Wilson (2002) found that the number of EDRs recorded from private pilots increased significantly during take-offs and landings of actual flights. These results demonstrate that EDA is influenced by cognitive activity and that it has utility as a measure of operator functional state.

4.1.5.4 Possibilities

SCR amplitudes may reflect the amount of affective or emotional arousal elicited by a stimulus or condition. Both amplitude and recovery of EDA have been demonstrated to be sensitive to certain aspects of central information processing (Boucsein, 1992) and may be used as indicators of mental strain. The frequency of spontaneous electrodermal changes (non-specific SCRs) is an indicator of emotional strain and has shown particular sensitivity during computerized work (Boucsein, 1992).

4.1.5.5 Limitations

EDA cannot define which particular functional state a human is experiencing.

4.1.5.6 General Advantages/Disadvantages of EDA

Compared to other biological signals taken from the skin, EDA can be regarded as a convenient measure for cognitive workload and the dynamics of emotional effort. EDA can be used as an additional indirect index of changes in the central nervous system to emotional and informational changes ("reaction to novelty"). It is a direct measure of sympathetic nervous system activity. EDR is associated with discrete environmental stimuli while other measures, such as heart rate, do not show individual responses to such events.

EDA is a slow system as demonstrated by the one-second to three-second response latency following an environmental event. Because of this lag, rapidly occurring changes in the environment cannot be followed with using EDR. As is the case with other electrophysiological measures, EDA is quite sensitive to external electromagnetic fields, movement of the operator, and thermal influences of heat or cold.

4.1.5.7 Apparatus Required

All of the EDA techniques require a readily available recording unit, electrodes, and electrolyte. The use of a special electrolyte, specific electrode site preparation, suitable amplifiers and electrodes has been recommended by some authors (Dawson et al., 2000). Portable recorders are available that can be used for ambulatory recording during work. Additionally, a PC computer with appropriate software is required for analyzing the data.

4.1.5.8 Personnel Required

The EDA technique requires the presence of skilled technicians. They need to test the recorders, fit operators with electrodes, monitor the data recording, and analyze the data.

4.1.5.9 Analysis Techniques

For SCL, the level of activity is measured with reference to a baseline period. The measure can be either the level during a given period of time or the number of spontaneous responses during a given time. Amplitude measures of SCR are made from the level prior to stimulation to the peak of the response. If several peaks appear in succession, then one of several methods of measurement must be selected. The latency of the SCR is recorded as the time from stimulus onset to the beginning of the SCR (Andreassi, 2000; Dawson et al., 2000; Stern et al., 2001). These measures can be accomplished with computer software but require trained personnel to review the data because of its variable nature.

4.1.5.10 References

Alajalova, N.A., Leonova, N.A., & Rusalov, V.M. (1975). About Parity of Features of Aim Formation at Accomplishment of Automated Mental Operations and of Super Slow Potentials of Brain. *Human Physiology*, *5*, 739-745.

Aldersons, A.A. (1985). *Electrodermal Response Mechanisms*. Riga: Znanie.

Andreassi, J.L. (2000). *Psychophysiology: Human Behavior and Physiological Response* (4th edition). Mahwah, NJ: Lawrence Erlbaum Associates.

Backs, R.W., & Boucsein, W. (2000). *Engineering Psychophysiology: Issues and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Backs, R.W., & Grings, W.W. (1985). Effects of UCS Probability on the Contingent Negative Variation and Electrodermal Response during Long ISI Conditioning. *Psychophysiology*, *22*, 268-275.

Boucsein, W. (1992). *Electrodermal Activity*. New York: Plenum Press.

Boucsein, W. (2000). The Use of Psychophysiology for Evaluating Stress-Strain Processes in Human-Computer Interaction (pp. 289-310). In R.W. Backs & W. Boucsein (Eds.), *Engineering Psychophysiology: Issues and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Boucsein, W., & Ottmann, W. (1996). Psychophysiological Stress Effects from the Combination of Night-Shift Work and Noise. *Biological Psychology*, *42*, 301-322.

Burov, A.Yu. (1986). The Evaluation of Functional State of Operators on Metrics of Mental Fitness for Work. *Human Physiology*, *12*, 281-288.

Dawson, M.E., Schell, A.M., & Filion, D.N. (2000). The Electrodermal System. In J.T. Cacioppo & L.G. Tassinary (Eds.), *Principles of Psychophysiology* (pp. 200-223). Cambridge: Cambridge University Press.

Heino, A., van der Molen, H.H., & Wilde, G.J.S. (1990). *Risk-Homeostatic Processes in Car Following Behaviour: Electrodermal Responses and Verbal Risk Estimates as Indicators of the Perceived Level of Risk during a Car-Driving Task*. Report from the University's Traffic Research Center. Groningen, The Netherlands: Groningen University.

Karpenko, A.V., Burov, A.Yu., Kalnish, V.V., Kapshuk, A.P., Grigorus, A.G., Grigorjantz, T.N., & Bobko, N.A. (1984). Control of Current Reliability of Operators Work on the Basis of Psychophysiological Criteria. *Nuclear Power Plant*, *7*, 156-162.

Kuhmann, W. (1989). Experimental Investigation of Stress-Inducing Properties of System Response Times. *Ergonomics*, *32*, 271-280.

Kuhmann, W., Boucsein, W., Schaefer, F., & Alexander, J. (1987). Experimental Investigation of Psychophysiological Stress Reactions Induced by Different System Response Times in Human-Computer Interaction. *Ergonomics*, *30*, 933-943.

Kuhmann, W., Schaefer, F., & Boucsein, W. (1990). Effekte von wartezeiten innerhalb einefacher aufgaben: Eine analogie zu wartezeiten in der mensch-computer-interaktion [Effects of Waiting Times within Sample Problems: An Analogy to Waiting Times in Human-Computer Interaction]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *37*, 242-265.

Lindholm, E., & Cheatham, C.M. (1983). Autonomic Activity and Workload during Learning of a Simulated Aircraft Carrier Landing Task. *Aviation, Space, and Environmental Medicine*, *54*, 435-439.

Lovibond, P.F. (1992). Tonic and Phasic Electrodermal Measurements of Human Aversive Conditioning with Long Duration Stimuli. *Psychophysiology*, *29*, 621.

Muter, P., Furedy, J.J., Vincent, A., & Pelcowitz, T. (1993). User-Hostile Systems and Patterns of Psychophysiological Activity. *Computers in Human Behavior*, *9*, 105-111.

Richter, P., Wagner, T., Heger, R., & Weise, G. (1998). Psychophysiological Analysis of Mental Load during Driving on Rural Roads – A Quasi-Experimental Field Study. *Ergonomics*, *41*, 593-609.

Schaefer, F., Kuhmann, W., Boucsein, W., & Alexander, J. (1986). Beanspruchung durch bildschirmtätigkeit bei experimentell variierten systemresponsezeiten [Stress Effects of Experimentally Varied System Response Times at Visual Display Units]. *Zeitschrift für Arbeitswissenschaft*, *40*, 31-38.

Stern, R.M., Ray, W.J., & Quigley, K.S. (2001). *Psychophysiological Recording* (2nd ed.). New York: Oxford University Press.

Wilson, G.F. (2002). An Analysis of Mental Workload in Pilots during Flight using Multiple Psychophysiological Measures. *International Journal of Aviation Psychology*, *12*, 3-18.

4.1.6 Electromyography (EMG)

4.1.6.1 Background

Electromyography (EMG) is a method of recording the electrical potentials originating in muscles over time. For a detailed description of EMG methodology, see Stern, Ray, and Davis (1980), from which the following technical description is largely abstracted. EMG recordings may be obtained by inserting electrodes into muscle tissue, or by attaching electrodes to the skin over the muscle or muscle groups of interest. At the cellular level, EMG reflects the spread of action potentials over skeletal muscle cells following neural (electrochemical) stimulation. Skeletal muscles are functionally divided into motor units that are distributed throughout a muscle and activated in unison. On an EMG record, this activation appears as a “wave”, the amplitude of which is a function of the number of muscle cells in the activated unit and their physical proximity to the electrodes.

Of particular importance, EMG is an especially sensitive indicator of the level of muscle activity related to variations in isometric contractions or “tension” (i.e., increases in muscle tonus that are not associated with actual movements and usually involve a single or relatively few motor units). Isotonic muscle activity (i.e., muscle contractions that result in skeletal movement) is more difficult to quantify because the gross muscle contraction alters the proximity of the electrodes to the source of the measured electrical potentials over time (i.e., over the duration of the movement). Also, the increased rate of firing and the increased number of motor units involved in actual movements tend to fuse to form a complex waveform that is not easily interpretable.

4.1.6.2 Rationale

EMG activity in small muscle groups, especially facial muscles such as the lateral frontalis and to a lesser extent the zygomaticus major and masseter, has been used as an indicator of anxiety and/or stress for many years. Its use in applications such as biofeedback therapy for the treatment of anxiety-related disorders is predicated on the notion that increased muscle tonus is part of the largely autonomic “fight or flight” response to stressors.

4.1.6.3 State of the Art

EMG recording requires AC amplifiers with high gain, high input impedance, and a frequency response range from 1-1000 Hz. Virtually any of the standard electrode types can be used. Modern amplifiers are available that are portable and digital, and off-the-shelf telemetry and computer hardware can be used for online monitoring and analysis of EMG signals. Modern electrodes are currently being developed that have substantial advantages over standard electrodes.

4.1.6.3.1 What EMG Can Tell Us

EMG signals have long been used as an index of stress, anxiety, and arousal. The relationship between EMG and psychological status continues to be explored, and most published reports do suggest a significant relationship. For example, Passchier and vd Helm-Hylkema (1981) showed that EMG amplitude is increased during imagination of a stressful situation versus a relaxing situation. Likewise, Carlson, Singelis, and Chemtob (1997) compared the effects of visually presented, unpleasant combat-related scenes versus neutral scenes in a group of combat veterans with Post Traumatic Stress Disorder (PTSD) versus a group of combat veterans without PTSD. They found more facial EMG activation (in frontalis, zygomaticus, and masseter) in the PTSD group when they viewed the unpleasant combat scenes. This result suggests that EMG may be useful for differentiating pathological from non-pathological stress reactions. Another recent study by Rissen, Melin, Sandsjo, Dohns, and Lundberg (2000) found that trapezius EMG activity was correlated with subjectively negative stress during repetitive work, although it was not correlated with positive emotional experiences, workload, or physical pain. This suggests that EMG measures may actually be useful for differentiation of "positive" from "negative" psychological states. It should be noted that facial, head, and neck muscle activity levels are intercorrelated, although they do not reflect the degree of tension in the general musculature (Graham et al., 1986). However, in addition to "stress," frontalis EMG activity increases in response to "concentration" (i.e., mental effort) during the performance of a mentally challenging task (Smith, Chung, & Berguer, 2000). Thus, during sedentary activities, EMG activity levels can reflect the extant level of stress, anxiety, and/or mental effort.

4.1.6.3.2 What EMG Cannot Tell Us

Although EMG activity appears to be reasonably sensitive under exquisitely controlled laboratory conditions, specificity is a problem. For example, EMG activity apparently increases in response to anxiety or stress levels, but not *only* in response to increased anxiety or stress. It also increases as a function of concentration or mental effort and, of course, as a function of actual movements. Additionally, tonic EMG levels have been reported to increase in response to sleep deprivation (Wilkinson, 1965), a finding which could be interpreted as suggesting that the act of resisting sleep onset requires increased concentration or mental effort. This also makes clear that EMG cannot be considered a straightforward indicator of "arousal" per se, since sleep deprivation clearly results in a state of reduced arousal characterized by relative brain deactivation (Thomas et al., 2000).

Also, the extent to which EMG measures may serve as a "scale" of anxiety and stress is unknown. No evidence has been found suggesting that varying levels of stress or anxiety produce corresponding variations in the level of EMG activation within individuals. Therefore, although EMG might serve as an indicator that the operator is in some way stressed, it would not, by itself, reflect the *amount or level of stress* experienced by the operator. That is, by itself EMG would not reflect the status of the operator with respect to his/her ability to perform duty-related tasks. At best, EMG might prove useful as one physiological measure within a suite of other measures and contextual information aimed at measuring and monitoring operator functional state but only for those operators engaged in sedentary activities, and even then, during periods of relative quiescence.

4.1.6.4 General Advantages/Disadvantages of EMG

EMG is typically obscured by movement artifacts, and is most effectively used with a subject at rest. It is good for detecting changes in muscle activity that do not result in observable movement. Any movement, even movements in limbs distal to the recording site, can potentially cause "movement artifact" that obscures the recording from the muscle or muscle group of interest.

Other potential artifacts include EEG (especially from the frontalis muscles) and EKG, as well as the artifacts typical to virtually all psychophysiological recordings (e.g., 60 Hz artifact that can be caused by loose electrodes and nearby electrical devices). Correct placement of leads is critical. The amplitude of the EMG signal depends upon the distance between the two electrodes from which the EMG measure is derived, and the longitudinal placement of the electrodes over the muscle or muscle group of interest. Thus, meaningful quantification of the EMG signal requires precise placement of electrodes.

4.1.6.5 Apparatus Required

An EMG or polygraph system with high gain, high input impedance, and a frequency response from 1-1000 Hz is required. A number of commercially available ambulatory recording systems are currently available that have this capability. Readily available electrodes can be used.

4.1.6.6 Personnel Required

Trained personnel are needed to precisely place and maintain the electrodes. Ambulatory recorders generally require replacement of batteries every 24 hours.

4.1.6.7 Analysis Techniques

Automatic scoring algorithms can be used to analyze the data for changes in EMG activity levels. On-line analyses require telemetry of EMG data from ambulatory operators.

4.1.6.8 References

Carlson, J.G., Singelis, T.M., & Chemtob, C.M. (1997). Facial EMG Responses to Combat-Related Visual Stimuli in Veterans with and without Post Traumatic Stress Disorder. *Applied Psychophysiology and Biofeedback*, *22*, 247-259.

Graham, C., Cook, M.R., Cohen, H.D., Gerkovich, M.M., Phelps, J.W., & Fotopoulos, S.S. (1986). Effects of Variation in Physical Effort on Frontalis EMG Activity. *Biofeedback and Self Regulation*, *11*, 135-141.

Passchier, J., & vd Helm-Hylkema, H. (1981). The Effect of Stress Imagery on Arousal and its Implications for Biofeedback of the Frontalis Muscles. *Biofeedback and Self Regulation*, *6*, 295-303.

Rissen, D., Melin, B., Sandsjo, L., Dohns, I., & Lundberg, U. (2000). *European Journal of Applied Physiology*, *8*, 215-222.

Smith, W.D., Chung, Y.H., & Berguer, R. (2000). A Virtual Instrument Ergonomics Workstation for Measuring the Mental Workload of Performing Video-Endoscopic Surgery. *Student Health Technology Information*, *70*, 309-315.

Stern, R.J., Ray, W.J., & Davis, C.M. (1980). *Psychophysiological Recording* (pp. 107-125). New York: Oxford University Press.

Thomas, M., Sing, H., Belenky, G., Holcomb, H., Mayberg, H., Dannals, R., Wagner, H., Thorne, D., Popp, K., Rowland, L., Welsh, A., Balwinski, S., & Redmond, D. (2000). Neural Basis of Alertness and Cognitive Performance Impairments during Sleepiness. I. Effects of 24 h of Sleep Deprivation on Waking Human Regional Brain Activity. *Journal of Sleep Research*, *9*, 335-352.

Wilkinson, R.T. (1965). Sleep Deprivation. In O.G. Edholm & A.L. Bacharach (Eds.). *The Physiology of Human Survival* (pp. 399-430). New York: Academic Press.

4.1.7 Eye Activity

4.1.7.1 Description of Eye Activity Data

The eyes are responsible for visual input to the brain, and monitoring eye-related activity has proven useful for understanding performance. Measures of eye activity include horizontal and vertical eye movements, blink activity, eye point of regard, and pupil diameter. Eye blinks have been shown to be related to fatigue and the visual demands of a task. Eye scan patterns are useful for determining how an operator views a scene such as a display panel. Pursuit eye movements are performed in order to maintain fixation (i.e., attention) on a moving target. The velocity of pursuit movements reaches 90°/s under optimal conditions (i.e., stimuli with large surfaces). Eye tracking has been used in several applications (e.g., in reading research, advertisement studies, studies investigating vestibular distractions by movements, and display design). The parameters of eye movement are customary physiological indicators of psychological constructs.

4.1.7.2 Background

4.1.7.2.1 As a Measure of Fatigue

Several eye activity parameters have been shown to be sensitive to time on task, which is linked indirectly to the onset of drowsiness in monotonous task environments. For example, using electrooculography (EOG) techniques, Stern and his colleagues (Stern, Boyer, & Schroeder, 1994; Stern, Walrath, & Goldstein, 1984) reported that blink duration and blink rate typically increase while blink amplitude decreases as a function of cumulative time on task. Others have found that saccade frequencies and velocities decline as time on task increases (McGregor & Stern, 1996; Schmidt, Abel, Dell'Osso, & Daroff, 1979). Morris and Miller (1996) demonstrated that during a 4.5-hr simulated flight, performance error rate increased, and blink amplitude, rate, long-closure rate, and saccade rate were the best predictors.

Overnight driving was associated with increased blink frequency while increased difficulty in the driving task (e.g., meeting oncoming traffic) resulted in decreased blink frequency (Summala, Häkkanen, Mikkola, & Sinkkonen, 1999). Torsvall and Akerstedt (1987) reported increased levels of slow eye movements in train drivers after driving about 4.5 hours. Wierwille, Wreggit, & Knipling (1994) reported that eyelid droop (percent time that the eyelid covers 80% or more of the pupil) may be useful for the determination of drowsiness during simulated driving tasks. This measure, PERCLOS, is gaining acceptance (Wierwille, 1999).

Using video analysis techniques, other investigators have shown that pupil diameter decreases as a function of subjective drowsiness (Lowenstein & Lowenfeld, 1962; Yoss, Moyer, & Hollenhorst, 1970). Concurrent with the tonic decrease in pupil size as a function of drowsiness is the occurrence of phasic oscillations that increase in amplitude (McLaren, Erie, & Brubaker, 1992).

4.1.7.2.2 As a Measure of Mental Workload and Attention

Blink rate has been found to decline with increased workload during a simulated air traffic control task (Brookings, Wilson, & Swain, 1996), in a flight simulator task (Veltman & Gaillard, 1998), and during actual flight (Wilson, Fullenkamp & Davis, 1994; Wilson, 2002). Blink closure duration declines with increasing mental workload, although the duration appears more related to visual demands than to general task complexity. Wilson, Fullenkamp, and Davis (1994) found that blink duration decreased to a greater extent during a taxing visual tracking task with minimal cognitive load than during a more cognitively challenging multiple task flight task. Paradoxically, blink rate can increase with additional workload. A study by Veltman and Gaillard (1998) demonstrated that combining a concurrent memory task with a flight control task resulted in elevated blink frequency compared to blink rates recorded during the flight task alone. They proposed that subvocal rehearsal during the memory task increased the likelihood that frontal lobe centers participating in eyelid control would be recruited. Tasks requiring eye movements, such as landing an aircraft, also increase blink rate (Hankins & Wilson, 1998; Wilson, 2002).

Hazardous roadway curves produced lowered driver blink rates (Richter, Wagner, Heger, & Weise, 1998). Blink closure duration was found to be sensitive to the cognitive demands of verbal versus digital communication formats in simulated helicopter flights (Sirevaag et al., 1993). Wilson (1992) reported total blink inhibition during a low-level airdrop maneuver in transport pilots.

Saccades, rapid movements of the eyes, are typically performed for locating an interesting target in the visual field. Therefore they can be used as an indicator of changes in attention. Saccade velocity interacts with saccade amplitude and the vigilance of the person and reaches maximum values of 100°/s to 520°/s. Saccade velocity cannot be influenced consciously and corresponds to operator state. The time between stimulus onset and the end of the saccade is the latency time. Average latency time is typically 150 to 250 ms. The time between the end of one saccade and the beginning of the following saccade is defined as the fixation or gaze duration. Increases in saccadic extent induced by higher workload levels are also associated with higher blink rates and greater blink durations (Fogarty & Stern, 1989; Hankins & Wilson, 1998). It is possible that the visual system takes advantage of the opportunity to blink during longer saccades, as is the case when a pilot is scanning for information both within and outside of a cockpit during landing. Thus, blink rate and duration have been described as more sensitive to visual than to cognitive task demands.

The effects of task workload on fixation dwell time, fixation frequency, and saccade extent have not been examined to the same extent as blink and pupil measures. Many flight simulation experiments confound cognitive and visual workload. For example, different flight maneuvers or situations of varying cognitive difficulty require very different kinds of visual activity (e.g., Hankins & Wilson, 1998; Itoh, Hayashi, Tsukui, & Saito, 1990; Katoh, 1997). Several visual search studies (e.g., Van Orden, Nugent, LaFleur, & Moncho, 1999; Walrath & Backs, 1989; Zelinsky, Rajesh, Hayhoe, & Ballard, 1997) have shown that more effortful search, as indicated by poorer performance accuracy and longer search times, is associated with increasingly greater numbers of fixations. Callan (1998) reported that the frequency of long fixations (exceeding 500 ms) correlated with the number of errors committed during a flight simulation task.

The pupil diameter controls the amount of light flowing into the eye. Pupil diameter, while also affected by changes in illumination, stimulus characteristics, and accommodative behaviors, has been shown to generally increase with higher cognitive processing levels (Backs & Walrath, 1992; Beatty & Wagner, 1978; Peavler, 1974). The diameter of the pupil is sensitive to changes in cognitive workload. Pupil changes can be dynamic, as during comprehension of discrete sentences (Just & Carpenter, 1993), or sustained, as is the case during digit span recall (Granholm, Asarnow, Sarkin, & Dykes, 1996). Granholm et al. (1996) reported that when cognitive resources were overtaxed, pupil diameter ceased increasing and began to decrease. This indicates that pupil diameter correlates nonlinearly with cognitive workload regardless of the specific visual demands imposed by a task.

4.1.7.2.3 Current State of the Art

Recent developments in video eye tracking systems, differential corneal reflection [CR]/pupil tracking in particular, provide measures of eye blink rate, blink duration, fixation duration and pupil diameter, as well as providing the operator's line of gaze (see Hudgins et al., 1998 for further information). Most systems enable the user to define scene planes in the environmental space that can be used to define the lines of gaze. This permits automatic analysis of scan patterns and fixation times in segments of the scene. As the operator looks at a certain portion of the scene, the line of gaze is registered and the duration of the fixation is calculated. This makes it possible to determine which display elements are perceived. With recording systems operating with higher time resolution, usually without video recording, it is possible to obtain saccade velocity.

These systems have become smaller and lighter so that head-mounted optics are comfortable for operators to wear at their workstations (Figure 13). Earlier systems did not permit head movements and could not be used in operational environments. Today's systems allow freer head movements by adding electromagnetic head trackers. If devices for recording or telemetry are used, the subject can even walk around.

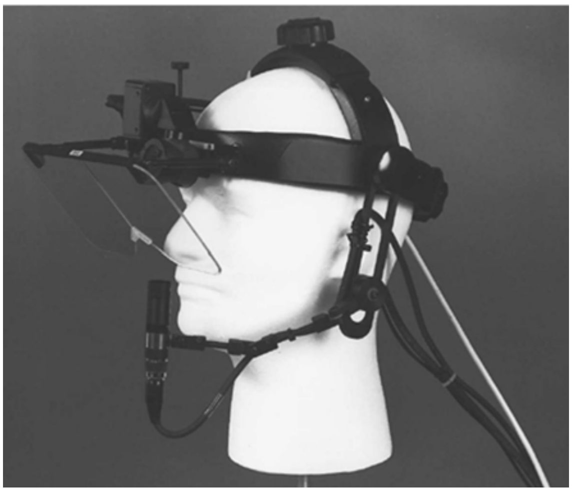

Figure 13: ASL 4000 – CR/Pupil Tracking Device with Head-Mounted Optics. (Adopted from the Website of Applied Science Laboratories Inc., http://www.a-s-l.com)

Another option is to use remote optics for the acquisition of eye movements without physical contact to the subject (Figure 14). Remote optics are typically placed in front of the operator so that they can detect the eye under usual head positions. They are especially useful if the area in which several pieces of information are displayed is limited and displays are arranged closely together (e.g. at workplaces within naval Command & Control Centers), so that only small head movements are necessary to read displays.

Figure 14: ASL 504 – Remote Optics. (Adopted from the Website of Applied Science Laboratories Inc., http://www.a-s-l.com)

Virtual reality (VR) technologies are being used in current research on system design for human-machine-interaction, training, and operational use. Experiments have been conducted in which pilots flew an aircraft with synthetic vision presented on a helmet mounted display (HMD). Eye tracking systems have been developed (Figure 15) which allow the measurement of eye movements with the operator wearing a head mounted display.

Figure 15: SMI iView – Head-Mounted Display with Integrated Eye Tracking Device.
(Adopted from the Website of Sensomotoric Instruments GmbH, http://www.smi.de)

4.1.7.2.4 As a Measure of Fatigue

Van Orden, Jung, and Makeig (2000) sought to determine the utility of eye activity measures and signal processing methods in a manner applicable to a real-time monitoring system. Five concurrent eye activity measures were used to model fatigue-related changes in performance during a visual compensatory tracking task. Mean tracking performance as a function of time across 18 sessions demonstrated a monotonic increase in error from 0 to 11 minutes, and a performance plateau thereafter. For each subject, moving estimates of blink duration and frequency, fixation dwell time and frequency, and mean pupil diameter were analyzed using non-linear regression and artificial neural network techniques. Models were derived using eye and performance data from one session and cross-validated on data from a second session. Correlation estimates to actual tracking performance averaged $\underline{R} = 0.68$. Mean cross-session correlations of estimated to actual tracking performance was $\underline{R} = 0.67$. Neural network models produced the lowest RMS error and highest correlation ($\underline{R} = 0.82$).

New analysis techniques are being applied to eye data to enhance the ability to detect and predict mental workload. Nonlinear regression analyses used blink frequency, fixation frequency, and pupil diameter to predict target density. Subject-specific artificial neural network models, developed through training on two or three sessions and subsequently tested on a *different* session from the same subject, correlated well with actual target density levels (Van Orden, Limbert, Makeig, & Jung, 2001). Blink rate and duration were among variables used to classify pilot workload using data from fighter aircraft flights (Wilson and Fisher, 1991) and to discriminate between two levels of task demand in a simulated aircraft landing task with an artificial neural network (Russell, Wilson & Monett, 1996). It may be possible to use eye activity, either alone or with other measures, to provide real-time assessment of operator functional state.

4.1.7.3 Limitations

4.1.7.3.1 What Eye Activity Data Can Tell Us

Because of the wide variety of parameters that can be measured or derived from eye activity their strengths and weakness will be considered individually (Galley, 2001).

The frequency of endogenous eye blinks has been found to be an indicator for the individual activation level. In the context of OFS assessment the relationship of eye blink rate with vigilance on the one hand

and visual complexity on the other is important. While performing tasks with complex information, blink rate decreases in order to ensure the perception of the information. With habituation or lowered interest, e.g. due to fatigue, blink rate increases. Blink amplitude has been found to be an indicator for emotional arousal. In connection with startle responses, blink amplitude reflects the intensity of the shock. Similar to blink rate, blink duration depends upon the visual complexity and the need of information perception.

Eye data provides several noteworthy advantages. The first involves the detection of drowsiness. EEG may be the only other measure that would provide any indication that an operator is becoming drowsy and at risk of performance decrement. The second benefit is derived from knowledge regarding the operator's attentional focus. While eye data cannot provide any indication of depth of focus or concentration, it can indicate which tasks in a multi-task, multiple-display work setting have *not* been attended to for some period of time. Such information would be very important to an adaptive system sensitive to the attentional demands and limits on an operator. Finally, eye tracking can be used for gross cursor control. In conducting usability and performance testing on prototype command and control consoles, Kellmeyer (2000) found that one master cursor is preferable to multiple cursors. However, relocating a cursor across several displays requires using either touch screens or significant trackball activity. It would be useful to have the cursor repositioned to the point of gaze under momentary control of an eye tracker via a trackball button. Accurate selection of display objects using an eye-controlled cursor is dependent upon the development of real-time eye position re-calibration methods.

Eye activity measures may prove useful for discriminating between mental workload levels. In a highly controlled choice reaction task, App and Debus (1998) found that saccadic velocity was dependent in part on subject arousal level. Saccades to targets were of higher velocity under more challenging conditions, and velocities declined as a function of time on task. The investigators surmise that saccadic velocities may provide some indication of operator stress. However, accurate measure of these velocities requires high temporal resolution, only recently available within video-based eye tracking systems. The extent to which saccadic velocity and other eye movement measures correlate with dynamic changes in pupil diameter and visual workload requires further investigation.

Saccade velocity is affected by habituation and fatigue. It is also useful as an indicator of activation especially in the field of drug induced vigilance degradations. Saccade latency time has been frequently used as a universal psychological metric comparable to manual reaction time. Fixation duration is commonly used as an indicator for the complexity of information perceived by the subject.

4.1.7.3.2 What Eye Activity Data Cannot Tell Us

Video-oculography depends upon the quality of the image of the eye, and factors that degrade image quality ultimately limit the utility of the method for determining functional state or attentional focus. Changes in ambient illumination and operator's use of corrective lenses or sunglasses have proven to be challenging obstacles in the development of prototype systems. Head and body movement is still a limitation for oculographic systems. Similar to other psychophysiological measures, optimal performance for the detection of drowsiness or changes in workload levels may require the development of individual models relating psychophysiological changes to behavioral measures. The development of these models may require the development of part-task simulations of the target task environment, and collection of operator data under a range of functional states. While Van Orden et al. (2001) have found reliable changes in eye activity for a broad range of workload levels, the identification of eye measures sensitive to subtle changes in workload and/or stress at nominally high workload levels remains a significant challenge. Traditional measures, such as blink frequency and duration, fixation frequency and duration, and pupil diameter, may interact in complex and nonlinear ways. New measures, such as saccadic velocity and long fixation frequency, require further rigorous testing to determine their sensitivity to subtle workload shifts.

4.1.7.4 General Advantages/Disadvantages of Eye Activity Data

Eye point-of-regard and EOG both provide information about operator's visual intake. A time history of eye scan behavior can be very valuable when evaluating visual displays. Most of these systems do not permit large head movements if calibrated data are required. EOG data can show the number and duration of blinks in many situations. This information can be used to evaluate the fatigue and the attentional state of an operator. Visual workload can also be estimated from these data.

Eye movements and the regulation of eye blinks and pupil diameter are the result of complex mechanisms with a large number of influencing factors such as activation level, drug consumption, task difficulty, visual environment, and lightning conditions. Therefore their measurement requires either carefully controlled conditions or knowledge of all relevant environmental parameters to ensure correct interpretation of psychological factors.

On the other hand video based measurement techniques provide useful information about the process of information perception in complex operational environments.

4.1.7.5 Apparatus Required

An oculometer is required for each operator in order to make eye point-of-regard measurements. These are usually worn on the head and held in position by a band or cap. Off head cameras are also used which are focused on the operator's face. Both types of systems typically consist of cameras to record eye position and the viewed scene. Most systems use infrared emitters to obtain information for point-of-regard. Head tracking devices are sometimes used in conjunction with the oculometer to permit more movement by the operator while still maintaining accurate eye position. A video recorder and a PC computer are required for data storage and analysis. Special software is required for calibrating the system, recording, and analyzing the data. Automatic analysis software is available to determine how often the eye fixated specific locations in the scene and how long the eye remained fixated on that location. Recently developed high-quality, remotely positioned face- and eye-tracking optical systems, with associated image processing algorithms, makes possible the unobtrusive and reliable acquisition of real-time eye activity for operator-state assessment and other purposes (see Pastoor, Lie, & Renault, 1999).

EOG measurement requires electrodes that are attached to the face above and below one eye for vertical movements and blinks and electrodes positioned at the external canthus of each eye to monitor horizontal movements. Electronic amplifiers are required to amplify and filter these data. To record saccadic activity, DC amplifiers are recommended. The data are digitized and processed with PC software. Software is used to detect blinks and measure the duration of the lid closure. Horizontal eye movements are also detected and measured using specialized software.

Several measurement techniques have been developed, but not all of these can be applied for all of the types of movements. Table 12 shows measurement techniques that have been widely applied and their most important properties.

4.1.7.6 Personnel Required

For eye point-of-regard procedures trained personnel are required to insure that the apparatus is properly fitted to the subjects, calibrated and that the data collection is valid. They also need to be familiar with the analysis techniques used. Personnel are needed to apply EOG electrodes, collect and process the data. This requires experience with these types of data and knowledge of the signals.

Table 12: Parameters for the Most Commonly Used Measurement Systems (based upon GALLEY, 2001).

Technique	EOG	IR-Limbus-Tracker	Direct Corneal-Reflectometry	Indirect Corneal-Reflectrometry	Purkinje Tracker
measurement principle	corneo-retinal potential	amount of light reflection	corneal reflex – pupil	corneal reflex	lense-reflex 1 vs. 4
accuracy	1°	<0.1°	0.5°	0.5°	0.01°
max. range, horiz.	±60°	±20°	±30°	±20°	±12°
max. range, vert.	±30°	±15°	±15°	±20°	±12°
signal-to-noise distance	0.3°-1.5°	<0.1°	0.3°	0.5°	<0.005°
head movements	no	no	yes	yes	no
scene video	no	no	yes	yes	no
typical time resolution [s^{-1}]	1000	100-165	50	50-600	1000
saccade amplitude fixation duration	yes	yes	yes	yes	yes
scan paths	no	no	yes	yes	yes
saccade velocity	yes	yes	no	+vdb	no
eye blink parameters	yes	yes	Rate only	Rate only	no
pupil diameter	no	no	yes	yes	no
costs	low	mid	high	high	high

4.1.7.7 Analysis Techniques

Automatic scoring software is available for analyzing eye point-of-regard data. With careful calibration and the use of a head tracking system data analysis can be performed on data collected in real world settings. Several commercial and academic packages have been developed. These permit determining eye point of regard over time and evaluation of how often regions were fixated and for how long. Software for analysis of pupil data and blink rates is usually included.

Automatic EOG scoring software has been developed that yields a count of eye blinks, the duration of each blink and a time history of the blinks. The time history can be used to compare with a timeline recorded during task performance.

4.1.7.8 References

App, E., & Debus, G. (1998). Saccadic Velocity and Activation: Development of a Diagnostic Tool for Assessing Energy Regulation. *Ergonomics*, *41*, 689-697.

Backs, R.W., & Walrath, L.C. (1992). Eye Movements and Pupillary Response Indices of Mental Workload during Visual Search of Symbolic Displays. *Applied Ergonomics*, *23*, 243-254.

Beatty, J., & Wagoner, B.L. (1978). Pupillometric Signs of Brain Activation Vary with Level of Cognitive Processing. *Science*, *199*, 1216-1218.

Brookings, J.B., Wilson, G.F., & Swain, C.R. (1996). Psychophysiological Responses to Changes in Workload during Simulated Air Traffic Control. *Biological Psychology*, *42*, 361-377.

Callan, D.J. (1998). Eye Movement Relationships to Excessive Performance Error in Aviation. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 1132-1136). Santa Monica, CA: Human Factors and Ergonomics Society.

Fogarty, C., & Stern, J.A. (1989). Eye Movements and Blinks: Their Relationship to Higher Cognitive Processes. *International Journal of Psychophysiology*, *8*, 35-42.

Galley, N. (2001). Physiologische grundlagen, meßmethoden und indikatorfunktionen der okulomotorischen aktivität. In: F. Rösler, *Grundlagen und anwendungen der psychophysiologie* (pp. 237-316). Göttingen, Germany: Hogrefe.

Granholm, E., Asarnow, R.F., Sarkin, A.J., & Dykes, K.L. (1996). Pupillary Responses Index Cognitive Resource Limitations. *Psychophysiology*, *3*, 457-461.

Hankins, T.C., & Wilson, G.F. (1998). A Comparison of Heart Rate, Eye Activity, EEG and Subjective Measures of Pilot Mental Workload during Flight. *Aviation, Space, and Environmental Medicine*, *69*, 360-367.

Hudgins, B., Leger, A., Dauchy, P., Pastor, D., Pongratz, H., Rood, G., South, A., Carr, K., Jarret, D., McMillan, G., Anderson, T., & Borah, J. (1998). *Alternative Control Technologies* (RTO-TR-7 AC/323(HFM)TP/3). Neuilly-sur-Seine, France: NATO-RTO.

Itoh, Y., Hayashi, Y., Tsukui, I., & Saito, S. (1990). The Ergonomic Evaluation of Eye Movements and Mental Workload in Aircraft Pilots. *Ergonomics*, *33*, 719-733.

Just, M.A., & Carpenter, P.A. (1993). The Intensity Dimension of Thought: Pupillometric Indices of Sentence Processing. *Canadian Journal of Experimental Psychology*, *47*, 310-339.

Katoh, Z. (1997). Saccade Amplitude as a Discriminator of Flight Types. *Aviation, Space, and Environmental Medicine*, *68*, 205-208.

Kellmeyer, D. (2000). Usability Testing & Analysis of Advanced Multimodal Watchstation Functions. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* (pp. 654-657). Santa Monica, CA: Human Factors and Ergonomics Society.

Lowenstein, O., & Lowenfeld, I.E. (1962). The Pupil. In H. Dawson (Ed.), *The Eye* (Vol. 3, pp. 231-267). New York: Academic Press.

McGregor, D.K, & Stern, J.A. (1996). Time on Task and Blink Effects on Saccade Duration. *Ergonomics*, *39*, 649-660.

McLaren, J.W., Erie, J.C., & Brubaker, R.F. (1992). Computerized Analysis of Pupillograms in Studies of Alertness. *Investigative Ophthalmology & Visual Science*, *33*, 671-676.

Morris, T.L., & Miller, J.C. (1996). Electrooculographic and Performance Indices of Fatigue during Simulated Flight. *Biological Psychology*, *42*, 343-360.

Pastoor, S., Liu, J., & Renault, S. (1999). An Experimental Multimedia System allowing 3-D Visualization and Eye-Controlled Interaction without User-Worn Devices. *IEEE Transactions on Multimedia*, *1*, 41-52.

Peavler, W.S. (1974). Individual Differences in Pupil Size and Performance. In M.P. Janisse (Ed.), *Pupillary Dynamics and Behavior* (pp. 159-175). New York: Plenum Press.

Richter, P., Wagner, T., Heger, R., & Weise. (1998). Psychophysiological Analysis of Mental Load during Driving on Rural Roads – A Quasi-Experimental Field Study. *Ergonomics*, *41*, 593-609.

Russell, C.A., Wilson, G.F., and Monett, C.T. (1996). Mental Workload Classification using a Backpropagation Neural Network. In: Dagli, C.H., Akay, M., Chen, C.L.P, Fernandez, B.R., & Ghosh, J. *Proceedings of the Artificial Neural Networks in Engineering (ANNIE'96) Conference*, *6*, 685-690.

Schmidt, D., Abel, L.A., Dell'Osso, L.F., & Daroff, R.B. (1979). Saccade Velocity Characteristics: Intrinsic Variability and Fatigue. *Aviation, Space, and Environmental Medicine*, *50*, 393-395.

Sirevaag, E.J., Kramer, A.F., Wickens, C.D., Reisweber, M., Strayer, D.L., & Grenell, J.F. (1993). Assessment of Pilot Performance and Mental Workload in Rotary Wing Aircraft. *Ergonomics*, *36*, 1121-1140.

Stern, J.A., Walrath, L.C., & Goldstein, R. (1984). The Endogenous Eyeblink. *Psychophysiology*, *21*, 22-33.

Stern, J.A., Boyer, D., & Schroeder, D. (1994). Blink Rate: A Possible Measure of Fatigue. *Human Factor*s, *36*, 285-297.

Summala, H., Häkkän, H., Mikkola, T., & Sinkkonen, J. (1999). Task Effects on Fatigue Symptoms in Overnight Driving. *Ergonomics*, *42*, 798-806.

Torsvall, L., & Akerstedt, T. (1987). Sleepiness on the Job: Continuously Measured EEG Changes in Train Drivers. *Electroencephalography and Clinical Neurophysiology*, *66*, 502-511.

Van Orden, K.F., Jung, T-P., & Makeig, S. (2000). Combined Eye Activity Measures Accurately Estimate Changes in Sustained Visual Task Performance. *Biological Psychology*, *52*, 221-240.

Van Orden, K.F., Limbert, W., Makeig, S., & Jung, T-P. (2001). Eye Activity Correlates of Workload during a Visuospatial Memory Task. *Human Factor*s, *43*, 111-121.

Van Orden, K.F., Nugent, W., Moncho, S., & Lafleur, B. (1999). *Assessment of Variable Coded Symbology using Visual Search Performance and Eye Fixation Measures.* (Technical Report No.: 99-4). San Diego, CA: Naval Health Research Center.

Veltman, J.A., & Gaillard, A.W.K. (1998). Physiological Workload Reactions to Increasing Levels of Task Difficulty. *Ergonomic*s, *41*, 656-669.

Walrath, L.C., & Backs, R.W. (1989). Time Stress Interacts with Coding, Density, and Search Type in Visual Search. *Proceedings of the Human Factors Society 33rd Annual Meeting* (pp. 1496-1500). Santa Monica, CA: Human Factors and Ergonomics Society.

Wierwille, W.W. (1999). Historical Perspective on Slow Eyelid Closure: Whence PERCLOS? Ocular Measures of Driver Alertness, *Technical Conference Proceedings, FHWA-MC-99-136* (pp. 31-53). Silver Spring, MD: Transportation Safety Associates.

Wierwille, W.W., Wreggit, S.S., & Knipling, R.R. (1994). Development of Improved Algorithms for On-Line Detection of Driver Drowsiness. *Convergence 94 Conference. Society of Automotive Engineers*, (pp. 331-340). Detroit, October, 1994.

Wilson, G.F. (1992). *Progress in the Psychophysiological Assessment of Workload*. (Technical Report No. AL-TR-1992-0007) Wright-Patterson AFB, OH: Armstrong Laboratory.

Wilson, G.F. (2002). An Analysis of Mental Workload in Pilots during Flight using Multiple Psychophysiological Measures. *International Journal of Aviation Psychology*, *12*, 3-18.

Wilson, G.F., & Fisher, F. (1991). The Use of Cardiac and Eye Blink Measures to Determine Flight Segment in F4 Crews. *Aviation, Space, and Environmental Medicine*, *62*, 959-961.

Wilson, G.F., Fullenkamp, B.S., & Davis, I. (1994). Evoked Potential, Cardiac, Blink, and Respiration Measures of Pilot Workload in Air-to-Ground Missions. *Aviation, Space, and Environmental Medicine*, *65*, 100-105.

Yoss, R.E., Moyer, N.J., & Hollenhorst, R.W. (1970). Pupil Size and Spontaneous Pupillary Waves Associated with Alertness, Drowsiness, and Sleep. *Neurology*, *20*, 545-554.

Zelinsky, G.J., Rajesh, P.N.R., Hayhoe, M.M., & Ballard, D.H. (1997). Eye Movements Reveal the Spatiotemporal Dynamics of Visual Search. *Psychological Scienc*e, *8*, 448-453.

4.1.8 Functional Magnetic Resonance Imaging (fMRI)

4.1.8.1 Background

The idea that blood flow is related to brain activity was first presented in 1890 by Charles S. Roy and Charles S. Sherrington (1890). Recent advances in technology have supported this hypothesis and are now able to provide images of the human brain in waking states while people are engaged in cognitive activity. The high spatial resolution of these images has been used to study cognitive activity in humans and permitted pin-pointing the engaged brain structures.

Earlier brain imaging techniques include Computer Aided Tomography (CAT). This technique uses X rays and provides images of the body that are related to the density of the various tissues. Positron Emission Tomography (PET) provides images of the brain's metabolic activity by measuring levels of radioactivity in the brain tissue. Radioactively labeled water is introduced into a vein in the arm and makes its way to the brain. Active brain areas absorb the radioactivity, which is then detected by the PET scanner. Only low doses of the radioactive label are necessary, but this is an invasive procedure. Changes in activity can be located to within a few millimeters with PET.

4.1.8.2 Magnetic Resonance Imaging (MRI) / Functional Magnetic Resonance Imaging (fMRI)

Magnetic Resonance Imaging (MRI) developed from work in the 1950s on a technique called nuclear magnetic resonance (NMR). This technique was used to investigate the chemical details of molecules and utilized the features that (1) a magnetic field can be manipulated to align atoms, and (2) radio waves can be used to perturb the atoms in a precise manner. The result of this manipulation and perturbation is the emission of detectable radio signals. The radio signals are also sensitive to the chemical environment of the atoms. The magnetic field and radio-wave pulses can be manipulated to reveal specific information about the sample of interest.

MRI was first used to provide anatomical information when it was discovered that NMR could form images by detecting protons, which are plentiful in human tissue. The images formed in this way were found to provide much better detail than either x-ray or CT images. The change in the name of the technique from NMR to MRI coincided with the development of the technique for clinical applications; the term "nuclear" gave the wrong impression for a technique that is both non-invasive and does not involve exposure to any radioactive material.

In 1990, Ogawa, Lee, Kay and Tank were the first to demonstrate that small magnetic changes due to functionally induced variation in blood oxygenation could be accurately mapped to specific brain structures using MRI. Development of techniques such as blood oxygenation level dependent (BOLD) contrast imaging led to *functional* MRI (fMRI) being used to investigate biological function. BOLD imaging utilizes the fact that changes in local blood flow in the brain (or other tissue of interest) causes an increase in the proportion of oxyhemoglobin compared to deoxyhemoglobin. The increase in blood oxygen levels affects the magnetic properties of the hemoglobin and these small changes are detected and are the source used to generate the BOLD contrast image.

4.1.8.2.1 Rationale for the Use of fMRI

Functional MRI is non-invasive because it utilizes the change in venous oxygen concentration that is a direct result of brain activity. No radioisotopes are required, so repeated measurements can be taken from an individual over an extended period.

fMRI can provide a spatial resolution of 1-2 mm, which is better than the resolution achievable with EEG. Both anatomical and functional data can be generated for each subject, making structural identification of active regions possible. Accurate localization of the anatomical source of activity is far more difficult to achieve with other methods such as EEG or magnetoencephalograph (MEG). fMRI can also show patterns of activation between structures and changes in activation with time-on-task or the learning of a task.

4.1.8.3 State of the Art

BOLD image contrast is the most commonly used fMRI technique. When transient local synaptic activation occurs, there is an increase in regional Cerebral Blood Flow (rCBF). This blood flow increase is greater than that needed to cover the rise in oxygen consumption. This results in greater oxygen saturation of the venous blood (Thulborn, 1998). The BOLD technique utilizes the increase in the oxyhemoglobin:deoxyhemoglobin ratio on the venous side of the local intra-vascular compartment of the tissue being sampled to generate the image contrast. The images generated in fMRI consist of a time series of scans, each producing intensity measurements for a large number of picture elements (pixels) or volume elements (voxels) that correspond to regions of interest (ROIs) used for analysis. The time series can be analyzed using different statistical tests (see Haxby, Courtney & Clark, 1998). Image acquisition can be precisely synchronized to external stimuli with good resolution. In principle, even transient changes in blood oxygenation can be measured, similar to event related potentials (Ogawa, Lee, Kay & Tank, 1990).

The images produced using fMRI are usually interpreted by means of comparisons. Experiments are designed to compare the images between just two tasks, usually task-induced activity and a baseline (subtraction), or between a series of tasks that have been varied systematically to examine different task levels or manipulations in more detail (correlation). Subtraction is the most commonly used experimental design, but correlation is, potentially, the more powerful technique since it provides the capability to look for quantitative relationships between a cognitive or behavioral parameter (for example attention, learning, or task difficulty) and neural activity.

Fast imaging methods such as echoplanar imaging (EPI; Turner & Jezzard, 1994) and spiral imaging allow for a volume of cross-sectional images to be constructed that can cover most or all of the brain in 2 to 6 seconds. This means that fMRI can be used to monitor the rate of change in the oxygen signal generated by functional changes in blood flow. It should be noted that this is not the same as being able to image the change in neuronal activity, which is much faster than the change in blood flow and still requires the use of techniques such as EEG or MEG, which have faster temporal resolution but poorer spatial resolution than fMRI.

4.1.8.3.1 What Can fMRI Tell Us?

Brain imaging allows the mental operations of a specific task or state to be linked to specific networks of brain areas that contribute to the execution of specific functions, or that renders visible conscious function (Posner & Raichle, 1997).

Error Detection

Event-related fMRI can be used to investigate the contribution of specific brain structures to performance. For example, Carter and colleagues (1998) determined how the anterior cingulate cortex (ACC) contributes to performance by detecting the conditions under which errors were likely to occur. Event-related potential studies showing error-related negativity with a probable medial frontal generator have led to the hypothesis that the ACC monitors and compensates for errors. fMRI techniques (spiral scanning) were able to confirm a specific performance monitoring role for the ACC. Further research showed that the ACC implements executive control through the on-line detection of response competition. Besides increased activation of the ACC when errors occur, the ACC is also active in tasks such as the Stroop test when error rates are low (Carter et al., 1998).

Attention

Functional MRI can be used to investigate the neural events that reflect attentional processes and changes in information processing. Functional MRI studies have indicated that the thalamus is involved in mediating the interaction of attention and arousal in humans (Portas, Rees, Howesman, Josephs, Turner, & Frith, 1998). Blindsight studies (e.g., Sahraie, Weiskrantz, Barbur, Simmons, Williams, & Brammer, 1997) have provided information on which brain structures are involved in conscious and unconscious processing of visual signals. These fMRI studies add to our knowledge about how the brain organizes visual processing and about the cognitive mechanisms that are used to make sense of the visual information.

4.1.8.3.2 What Can fMRI Not Tell Us?

Functional MRI does not provide information better than a few millimeters resolution, so it is not possible to use this technique to directly look at the cellular level. It is possible, however, to coordinate human studies (fMRI) and animal studies (at the cellular level) to build a better understanding of how mental activity is generated at the cellular level (Desimone & Duncan, 1995).

Brain response, in terms of underlying electrochemical events, take tens of milliseconds, whereas the responses detected by fMRI (changes in cerebral blood flow) occur over hundreds of milliseconds. This leads to limited temporal resolution for fMRI of thousands of milliseconds (Norris & Wise, 2000). The hemodynamic changes measured by fMRI do not reach their maximum for 4 to 8 seconds, whereas changes in neural activity triggered by sensory stimulation or motor activity occur on the order of tens or hundreds of milliseconds. Other techniques such as EEG or MEG can provide better temporal resolution. Gevins, Brickett, Reutter, and Desmond (1991) have demonstrated how MRI can be used to improve the quality of EEG data. Spatial information from the MRI was used to develop spatial signal enhancing techniques to correct blur distortion when using EEG. Spatial detail approaching that of O15 PET is claimed for the EEG using these techniques.

Movement artifacts can limit experimental design, since subject movement needs to be restricted. This is usually accomplished by the use of some type of head clamp. Movement artifacts can change the contrast of the BOLD image and also make it difficult to average over repeated trials. The scanning process produces auditory noise and some researchers opt to use active noise cancellation techniques.

The large and expensive apparatus and the need for advanced analysis techniques make this an expensive and highly skilled process. Current systems do not lend themselves to in-field or real-time applications.

Optical imaging techniques such as near-infrared spectroscopy (NIRS) may lead to the development of portable systems with some of the properties of fMRI since they also use regional blood flow to determine activity.

4.1.8.4 References

Carter, C.S., Braver, T., Barch, D.M., Botvinick, M., Noll, D., & Cohen, J.D. (1998). Anterior Cingulate Cortex, Error Detection and Online Monitoring of Performance. *Science*, *280*, 747-749.

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscienc*e *18*, 193-222.

Gevins, A., Le, J., Brickett, P., Reutter, B., & Desmond, J.E. (1991). Seeing Through the Skull: Advanced EEGs use MRIs to Accurately Measure Cortical Activity from the Scalp. *Brain Topography*, *4*, 125-131.

Haxby, J.V., Courtney, S.C., & Clark, V.P. (1998). Functional Magnetic Resonance Imaging and the Study of Attention. In: R. Parasuraman (Ed.), *The Attentive Brain (*pp. 123-142). Cambridge: MIT Press.

Ogawa, S., Lee, T., Kay, A., & Tank, D. (1990). Brain Magnetic Resonance Imaging with Contrast Dependent on Blood Oxygenation. *Proceedings of the National Academy of Science USA, 87*, 9868-9872.

Norris, D., & Wise, R. (2000). The Study of Prelexical and Lexical Processes in Comprehension: Psycholinguistics and Functional Brain Imaging. In, M.S. Gazziniga (Ed.), *The New Cognitive Neurosciences* (pp. 867-880). Cambridge: MIT Press.

Portas, C.M., Rees, G., Howesman, A.M., Josephs, O., Turner, R., & Frith, C.D.A. (1998). Specific Role for the Thalamus in Mediating the Interaction of Attention and Arousal in Humans. *The Journal of Neuroscience*, *18*, 8979-8989.

Posner, M., & Raichle, M. (1997). *Images of Mind*. New York: Scientific American Library.

Roy, C., & Sherrington, C. (1890) On the Regulation of Blood Supply of the Brain. *Journal of Physiology (London) 11*, 85-108

Sahraie, L., Weiskrantz, J.L., Barbur, J.L., Simmons, A., Williams, S.C.R., & Brammer, M.J. (1997). Pattern of Neuronal Activity associated with Conscious and Unconscious Processing of Visual Signals. *Proceedings of the National Academy of Science USA, 94*, 9406-9411. Neurobiology.

Thulborn, K.R. (1998). A BOLD Move for fMRI. *Nature Medici*ne, *4*, 155-156.

Turner, R., & Jezzard, P. (1994). Magnetic Resonance Studies of Brain Functional Activation using Echo-Planar Imaging. In R.W. Thatcher, M. Hallett, T. Zeffiro, E.R. John, & M. Huerta (Eds.), *Functional Neuroimaging: Technical Foundations* (pp. 69-78). San Diego, CA: Academic Press.

4.1.9 Near-Infrared Spectroscopy

4.1.9.1 Description of Near-Infrared Spectroscopy

Near-infrared Spectroscopy (NIRS) is a technique for analyzing the chemical composition of a material from the absorption spectrum of red and near-infrared light. Light at these wavelengths easily penetrates biological tissues, including bone, and the extinction of light by water (the largest component of most biological tissues) is low in the 700 to 1000 nm range. Further, the absorption spectra of the oxygen carrying pigment hemoglobin in the blood and cytochome c oxidase (Cyt_{aa3}) in cell mitochrondria are well defined. Changes in the absorption spectrum of both chromophores in the cerebral tissue, and thus the amount of oxygenated hemoglobin (HbO_2), de-oxygenated hemoglobin (Hb), and oxidized Cyt_{aa3} can be measured with a time resolution of less than ½ second (Jobsis, 1977). From these metrics, estimates of the localized changes in cerebral blood flow, cerebral oxygen utilization, and the oxidative state of the cerebral tissue can be monitored in real-time.

4.1.9.2 Background

Real-time cerebral NIRS can provide a non-invasive, real-time, accurate metric of the severity of the hypoxic state of the operator due to low-oxygen levels in the environment, decreased cerebral blood flow (Ferrari, Zanette, Giannini, Sideri, & Fiesch, 1987; Ferrari, Zanette, Giannini, Sideri & Fieschi, 1986, Germon, Kane, Manara & Nelson, 1994, Kuroda, Houkin, Abe, Hoshi & Tamura, 1996), sleep apnea, intense exercise, or carbon monoxide exposure. Thus NIRS can be an effective monitor of the severity of the environmental stress imposed on the cognitive capabilities of the brain. In addition, changes in the activity of the cerebral cortex due to cognitive processing is believed to be reflected in changes in local cerebral blood flow and tissue oxygenation. Such changes correlate with EEG activity recorded from the overlying scalp, PET (Hock *et al*., 1997) and fMRI scans, as well as cognitive performance metrics.

4.1.9.3 State of the Art

The non-invasive transmission of near-infrared light into cerebral tissue is done with either light-emitting diodes (LEDs) or laser diodes using *optrodes* placed directly on the scalp. Reflected light from the cerebral tissue is received and amplified with photodiodes sensitive to the wavelengths of interest. Initial efforts in non-invasive monitoring of brain oxygen state required multiple racks of electrical signal conditioning equipment (Jobsis, 1977). However, the desire to develop the technique for the neurosurgical community has dramatically decreased the size and cost of the technologies. The costs of the sensor materials is decreasing, especially with the development of newer light-emitting diodes that can replace the laser technologies. The signal conditioning components of the both the emitter and detection circuits tend to be the most expensive part of the systems. The size and weight of overall system hardware has been reduced resulting in the use of experimental systems in high-performance aircraft (Kobayashi & Miyamoto, 2000) and human centrifuge studies (Fraser, Shender, Forster & Hrebien., 2000).

Enhancements to non-invasive *in vivo* NIRS techniques are on-going, including the development of continuous intensity spectroscopy, time-resolved spectroscopy, and intensity or frequency modulated spectroscopy. Multi-sensor systems have been developed allowing for computerized optical density, optical coherence, and other forms of optical topographical imaging of the brain state and structure to be performed in real-time, and in conjunction with multi-channel EEG monitoring.

4.1.9.3.1 What Near-Infrared Spectroscopy Can Tell Us

NIRS can measure changes in cerebral blood flow, which likely reflects a change in metabolic demand by the cerebral neurons. Thus these changes reflect functional activation of various sites within the cerebral cortex. Given the ability of NIRS to quantify the delivery of oxygenated blood to different areas of the brain, and monitor the actual redox state of the cerebral mitochondria, it is the most direct physiological measure of brain functional state.

Most of the work correlating brain function with NIRS has focused on changes in HbO_2 and Hb during activation of various brain centers. In general, activation of brain cells increases local cerebral blood flow out of proportion to oxygen metabolism (Fox & Raichel, 1986). This activation of *single* centers of the brain can be demonstrated with high temporal resolution. During cognitive tasks and visual simulation there are increases in cerebral tissue HbO_2, often with corresponding decreases in Hb over the corresponding cortical areas (Heekeren et al., 1997; Villringer, et al., 1994, Villringer, et al., 1997; Meek, et al., 1995).

Hirth, Obrig, Valdueza, Dirnagl and Villringer, (1997) demonstrated increases in HbO_2 and decreases in Hb using optrodes placed over the left motor cortex during right hand movements but not with left hand movements. Simultaneous spectroscopy of both hemispheres can provide information on the asymmetry of cerebral activation during mental stimulus (Tamura, Hoshi & Okada, 1997). NIRS has also had application in psychiatric evaluation, demonstrating clear differences between cerebral function of schizophrenic and normal controls; and between patients with Alzheimer's dementia and healthy elderly subjects (Okada, Tokumitsu, Hoshi, & Tamura, 1994; Hock *et al*., 1997).

While there is strong evidence for changes in HbO_2 and Hb during functional activation of various parts of the cerebral cortex, recent work by Villringer *et al*. (1997) have shown statistically significant increases in Cyt_{aa3} oxidation during visual and motor activation tasks.

4.1.9.3.2 What Near-Infrared Spectroscopy Cannot Tell Us

Like all physiological metrics, NIRS does not directly measure the cognitive state. Since it measures changes in cerebral oxygen state due to both changes in oxygen delivery to the brain and changes in cerebral tissue activation as a result of cognitive processing, it does not necessarily provide an accurate representation of either during cortical functioning under hypoxic stress. Systemic decreases in cerebral oxygen delivery to the brain should be reflected in a decreased HbO_2 and reduced Cyt_{aa3} over a wide areas of cortical tissue. Specific activation of cortical centers will result in regional changes in these variables, multi-channel NIRS sensors distributed over the cerebral cortex will allow simultaneous separation and quantification of both hypoxic induced and functional induced changes in brain state.

As in the case of other metrics, such as PET and fMRI, the correlation between active brain function and operator functional state and changes in the absorption spectra of oxygenated chromophores in cerebral tissue is not necessarily causal.

4.1.9.4 General Advantages/Disadvantages of Near-Infrared Spectroscopy

Both the NIRS technology and its application to operator functional state estimation are relatively new fields, and thus do not have the general acceptance accorded more mature techniques. The cost of laboratory and field portable hardware is still on the order of one to two magnitudes greater than EEG and ECG monitoring systems. The major limitation is the lack of research into the capabilities of the technology, especially research into the correlation between information collected with various NIRS techniques, performance tasks, and other metrics of operator functional state, such as EEG and eye-movement.

Given the falling costs of the technology, the availability of commercial products, and the development of optical tomography there are extensive opportunities to undertake research in the applicability of the technique to OFS assessment.

The redox state of Cyt_{aa3} is the most direct measure of cerebral function and should be the most specific measure of both hypoxic stress and cerebral tissue activation. However, changes in the absorption spectra of Cyt_{aa3} are very small, such that the total attenuation of light displayed by hemoglobin is ten-fold larger than the cytochrome chromophore. Thus there is potential for significant "cross-talk" and contamination of the Cyt_{aa3} spectrum. Major research efforts are required to address these issues.

4.1.9.5 Apparatus Required

Only two commercial NIRS instruments are available at this time although a number of published studies use custom designed hardware. The NIRS-300 from Hammarstu, Japan is the only commercial device providing information on the oxidative status of Hb and Cyt_{aa3}.

4.1.9.6 Personnel Required

Operation of the commercial NIRS equipment and the attachment of the "optrodes" is no more difficult than using commercial EEG and ECG hardware. Any computerised data acquisition system can be used to digitise and store the analog signals corresponding to the level of the output parameters.

4.1.9.7 Analysis Techniques

Depending on the design, NIRS instrumentation produces time series data for Hb, HbO2, total tissue oxygen, total blood volume, the levels of reduced and oxidized cytochrome or various ratios of these parameters. Any analysis software capable of analyzing multiple time series data can be used in monitoring OFS information provided by the NIRS technology.

4.1.9.8 References

Ferrari, M., Zanette, E., Giannini, I., Sideri, G., & Fieschi, C. (1986). Effects of Carotid Artery Compression Test on Regional Cerebral Blood Volume, Hemoglobin Oxygen Saturation, and Cytochrome-c-oxidase Redox Level in Cerebrovascular Patients. *Advances in Experimental Medicine and Biology*, *200*, 213-221.

Ferrari, M., Zanette, E., Giannini, I., Sideri, G., & Fiesch, C. (1987). Effects of Carotid Compression, as Assessed by Near Infrared Spectroscopy, upon Cerebral Blood Volume and Hemoglobin Oxygen Saturation. *Journal of the Royal Society of Medicine*, *80*, 83-87.

Fox, P.T., & Raiche, M.E. (1986). Focal Physiological Uncoupling of Cerebral Blood Flow and Oxidative Metabolism during Somatosensory Stimulation in Human Subjects. *Neurobiology*, *83*, 1140-1144.

Fraser, W.D., Shender, B.S., Forster, E.M., & Hrebien, L. (2000). Cerebral Tissue Oxygenation and Pulse Wave Delay (PWD) during Negative to High Positive Gz Transitions. *Aviation. Space Environmental. Medicine*, *71*, 275.

Germon, T.J., Kane, N.M., Manara, A.R., & Nelson, R.J. (1994). Near-Infrared Spectroscopy in Adults: Effects of Extracranial Ischaemia and Intracranial Hypoxia on Estimation of Cerebral Oxygenation. *British Journal of Anesthesia*, *73*, 503-506.

Heekeren, H.R., Obrig, H., Wenzel, R., Eberle, K., Ruben, J., Villringer, K., Kurth, R., & Villringer, A. (1997). Cerebral Hemoglobin Oxygenation during Sustained Visual Simulation – A Near-Infrared Spectroscopy Study. *Philosophical Transactions of the Royal Society, London B. Biological Sciences*, *3345*, 825-828.

Hirth, C., Obrig, H., Valdueza, J., Dirnagl, U., & Villringer, A. (1997). Simultaneous Assessment of Cerebral Oxygenation and Hemodynamics during Motor Task. A Combined Near-Infrared and Transcranial Doppler Sonography Study. *Advances in Experimental Medicine & Biology, 411*, 461-469.

Hock, C., Villringer, K., Muller-Spahn, F., Wenzel, R., Heekeren, H., Schuh-Hofer, S., Homann, M., Minoshima, S., Schwaiger, M., Dirnagl, U., & Villringer, A. (1997). Decrease in Parietal Cerebral Oxygenation during Performance of a Verbal Fluency Task in Patients with Alzheimer's Disease Monitored by Means of Near-Infrared Spectroscopy (NIRS) – Correlated with Simultaneous rCBF-PET Measurements. *Brain Research*, *755*, 293-305.

Jobsis, F.F. (1977). Non-Invasive, Infrared Monitoring of Cerebral and Myocardial Oxygen Sufficiency and Circulatory Parameters. *Science*, *198*, 1264-1267.

Kobayashi, A., & Miyamoto, Y. (2000). In-Flight Cerebral Oxygen Status: Continuous Monitoring by Near-Infrared Spectroscopy. *Aviation, Space, and Environmental Medicine*, *71*, 177-183.

Kuroda, S., Houkin, K., Abe, H., Hoshi, Y., & Tamura, M. (1996). Cerebral Hemodynamic Changes during Carotid Artery Balloon Occlusion Monitored by Near-Infrared Spectroscopy. *Neurologia Clinica Chirurgica (Tokyo), 36*, 78-86.

Meek, J.H., Elwell, C.E., Khan, M.J., Romaya, J., Wyatt, S., Delpy, D.T., & Zeki, S. (1995). Regional Changes in Cerebral Hemodynamics as a Result of a Visual Stimulus Measured by Near-Infrared Spectroscopy. *Proceedings of the Royal Society, London B Biological Science*, *261*, 351-356.

Okada, F., Tokumitsu, Y., Hoshi, Y., & Tamura, M. (1994). Impaired Interhemispheric Integration in Brain Oxygenation and Hemodynamics in Schizophrenia*. European Archives of Psychiatry and Clinical Neuroscience*, *244*, 17-25.

Tamura, M., Hoshi, Y., & Okada, F. (1997). Localized Near-Infrared Spectroscopy and Functional Optical Imaging of Brain Activity. *Transactions of the Royal Society of London, B Biological Sciences*, *352*, 737-742.

Villringer, A., Planck, J., Stodieck, S., Botzel, K., Schleinkofer, L., & Dirnagl, U. (1994). Non-Invasive Assessment of Cerebral Hemodynamics and Tissue Oxygenation during Activation of Brain Cell Function in Human Adults using Near-Infrared Spectroscopy. *Advances in Experimental Biology and Medicine*, *345*, 559-565.

Villringer, K., Minoshima, S., Hock, C., Obrig, H., Ziegler, S., Dirnagl, U., Scwaiger, M., & Villringer, A.A. (1997). Simultaneous PET and Near-Infrared Spectroscopy Study. *Advances in Experimental Medicine and Biology*, *13*, 303-307.

4.1.10 Oximetry

4.1.10.1 Description of Oximetry

Pulse oximetry is a simple non-invasive method of monitoring the percentage of haemoglobin (Hb) which is saturated with oxygen. It provides estimates of arterial oxyhemoglobin saturation (SaO2) by utilizing selected wavelengths of light to non-invasively determine the saturation of oxyhemoglobin (SpO2). The pulse oximeter consists of a probe attached to the person's finger or ear lobe which is linked to a computerised unit . The unit displays the percentage of Hb saturated with oxygen together with an audible signal for each pulse beat, a calculated heart rate and in some models, a graphical display of the blood flow past the probe.

4.1.10.2 Background

Current pulse oximeters measure the differential absorption of two wavelengths (colours) of light projected through the finger or other tissue. It is based on two physical principles: different colours of light are absorbed differently by oxygenated hemoglobin and deoxygenated hemoglobin; and the fluctuating volume of arterial blood between the source and detector which adds a pulsatile component to the absorption (Severinghaus and Kelleher, 1992). Tissue, bone and venous blood absorb a relatively constant amount of light, producing unknown but constant background absorption. Each time the heart beats, a pulse of arterial blood flows to the tissue. The influx of blood increases the absorption at both wavelengths. The ratio of absorption at these two wavelengths varies with the oxygen saturation. This ratio is converted to SpO_2 via empirical tables or calibration curves derived from volunteer desaturation studies.

4.1.10.3 State of the Art

4.1.10.3.1 Oxygen and Hemoglobin

In 1940 Glen Millikin developed a lightweight oximeter to help the military solve their aviation hypoxia problem. Oximetry measures the percentage of hemoglobin saturated with oxygen by passing specific wavelengths of light through the blood.

The hemoglobin molecule consists of 10,000 atoms, four of which are iron atoms that attract and hold oxygen. Each red blood cell contains about 250 million hemoglobin molecules. There are approximately 5,000 cc's of blood in the average individual and each cc contains five billion red blood cells. When oxygen is bound to the hemoglobin, it is called oxyhemoglobin.

Oxygen is a clear, odorless gas that accounts for 21% of the gases in the air around us. It is essential for the process our body uses to produce the energy needed for metabolism. Too much or too little oxygen (hypoxia) can cause illness or death; therefore, it is necessary to be able to quantify the amount of oxygen in the blood.

Oxygen can be measured in two forms: partial atmospheric pressure of oxygen (PaO2) and oxygen saturation (SaO2).

When the PaO2 is determined, we are measuring the actual amount of oxygen that is dissolved in the blood. Since the pressure of 1 atmosphere is 760 mm of Hg and oxygen comprises 21% of the atmosphere at sea level, we find 21% of 760 which is 160. After adjusting for dead airway space, elevation, subject's temperature, and water vapor, the range of a normal PaO2 should be between 90-106 mm of Hg. There is a relationship between the amount of oxygen dissolved in the blood and the amount attached to the hemoglobin.

As the pressure of oxygen increases, the hemoglobin saturation increases. A pressure of 105 or above will completely saturate the hemoglobin. More oxygen can still be diffused into the blood but the hemoglobin is at its maximum capacity. By using the pulse oximeter we can indirectly assess the PaO2 by measuring the SpO2. For example: 97% saturation = 97 PaO2 *(normal*); 90% saturation = 60 PaO2 (*danger*), and 80% saturation = 45 PaO2 (*severe hypoxia*).

The functions of hemoglobin are oxygen pickup and delivery. The hemoglobin's affinity can be increased or decreased due to various situations. If hemoglobin has an increased affinity, it is highly saturated; but oxygen is less available for release to the tissues due to the strong bond. The reverse is also true. Shifts occur due to an alteration in normal pH, CO2 levels, temperature, and 2-3- diphosphoglycerate. 2-3- diphosphoglycerate is a normal product of red blood cell metabolism. An increase in 2-3- diphosphoglycerate can be caused by residence at high altitude, anemia, chronic hypoxemia, chronic alkalosis A decrease in 2-3- diphosphoglycerate can be caused by infusion of stored bank blood, hypophosphatemia, chronic acidosis.

An increase in hemoglobin's affinity for oxygen can be caused by alkalosis, decreased PaCO2, hypothermia, or decreased 2-3- diphosphoglycerate. During these circumstances, a pulse oximeter reading of 95%, which is usually considered as normal, denotes a PaO2 of 76 showing that the subject is hypoxic.

A decrease in the hemoglobin's affinity for oxygen can be caused by acidosis, increased PaCO2, fever, or increased 2-3- diphosphoglycerate. When such an event occurs, a SaO2 of 75% (usually considered severe hypoxia) denotes a PaO2 of 88. This subject is not nearly as hypoxic as the SaO2 would lead us to believe.

The normal SpO2 value for adults with no lung disease or smokers is saturation greater than 95%.

4.1.10.3.2 Hypoxia Measurement

Hypoxia is defined as a lack of oxygen to the tissues sufficient to cause impairment of function. Traditionally, there are four different mechanisms described which can lead to hypoxia (Sheffield and Heimbach, 1996). Hypoxic hypoxia is a problem of oxygen diffusion from lungs to blood, caused in aviation by a reduced driving pressure of oxygen in the atmosphere and lungs. Hypaemic hypoxia is an oxygen transport problem, typically caused by insufficient haemoglobin. Stagnant hypoxia is a blood flow problem, as may be seen in a patient with cardiac failure, or blood vessel obstructions. Finally, histotoxic hypoxia refers to hypoxia in the mitochondria themselves, where the cells are unable to utilise oxygen perhaps due to the presence of cellular toxins.

Exposure to high altitude can produce hypoxic hypoxia can occur. The fractional composition of the atmosphere remains relatively constant up to an altitude of 300 000 feet, but it is the decreasing partial pressure of oxygen in the atmosphere and the alveolus that results in hypoxic hypoxia. As oxygen comprises approximately 21% of the air, at sea level the partial pressure of oxygen is 21% of 760 mmHg, or 160 mmHg. At an altitude of 8000 feet, which is a common cabin altitude in pressurised aircraft, the partial pressure of oxygen in cabin air is 21% of 565 mmHg, which gives 118 mmHg. Thus, partial pressures of oxygen in ambient air, alveolar air and arterial blood decrease at high altitudes. Above 10,000 feet, saturations drop steeply as PAO2 falls below about 55 mmHg. Cabin pressurisation systems aim to keep altitude below 10,000 feet, well within what is known as the “Physiological Zone”, where normal individuals at rest will experience little in the way of symptoms (Ernsting, 1984).

Above 18,000 feet, falling barometric pressure has another potentially dangerous effect – that of decompression illness (Murrison and Francis, 1991). Decompression illness is due to the evolution of nitrogen bubbles out of solution in body fluids as the partial pressure of nitrogen in the air around us decreases. This states that the partial pressure of a gas in solution is proportional to the partial pressure of that gas above the solution, with which it does not combine chemically. The longer a person spends at altitudes above 18,000 feet, the more likely it is that decompression illness will occur. Pressurisation systems minimise the risk of this problem by limiting the pressure environment to below 10,000 feet.

At sea level the oxygen saturation is typically 95% to 100%. At 8,000 feet, the normal oxygen saturation drops to about 90% and continues to decrease as one goes to higher altitudes. As O2 sat decreases below 80%, measurable impairment of cognitive and physical performance begins. Those changes don’t occur immediately, but vary with the speed of ascent and the duration of exposure (Nehrenz, 1997).

Oxygen should therefore be delivered at altitude, or cabin pressurisation adjusted, to maintain arterial oxygen. Oxygenation at altitude can be predicted at sea level if arterial blood gases are known. Haemoglobin should be checked in all subjects prior to flight, as their oxygen carrying capacity may be severely limited, requiring oxygen supplementation.

At higher altitudes, blood oxygen saturation decreases, since there is less oxygen available in the air. Hypoxia, defined as a deficiency of oxygen reaching the tissues of the body, is now a risk. When oxygen saturation levels drop, bad things happen that are rarely perceived by the victim (at least in the early stages). Visual symptoms occur, including “tunnel vision” and a marked decrease in night vision. Other common symptoms of hypoxia include headaches, anxiety, panic sensation, inability to perform mathematical problems accurately, inability to program equipment such as a GPS, dizziness, nausea, headache, and confusion. Symptoms are different for each person, and can occur at altitudes far lower than most people would predict. Symptoms of hypoxia generally remain consistent for a particular person, but the altitude at which the onset of impairment occurs is highly variable from day to day.

For pilots, the most dramatic effect is reduced mental proficiency, dulling judgment, damaging memory and limiting the performance of discrete motor movements. Vision can be impaired, especially at night. For young pilots in top physical condition, the effects of less oxygen usually kick in at 12,500 feet.

Health and lifestyle factors like drinking, smoking, age and weight gain can amplify the effects of reduced oxygen content, revealing symptoms of hypoxia at altitudes as low as 4,000 feet.

Hypoxia is insidious and highly physiologically variable from pilot to pilot and in the same pilot from day to day. The requirements for supplemental oxygen use may be too liberal or too conservative, and there is no objective way for a pilot to know without using a pulse oximeter to measure the actual level of oxygen saturation. An overweight, out-of-shape, middle-aged smoker will become hypoxic at a far lower altitude than a young, athletic non-smoker.

One of the earliest physiological effects of hypoxia is a change in respiration from steady to cyclical. This change in involuntary breathing patterns interfere with respiratory efficiency and exacerbate the hypoxic effect of high-altitude flight.

The availability of small, low-cost pulse oximeters suitable for use in the cockpit provides an enormous leap forward in detecting and dealing with in-flight hypoxia. Although not perfect, the pulse oximeter which can be worn on a fingertip by both pilot and passengers gives an almost instantaneous oxygen saturation reading.

4.1.10.4 Limitations

4.1.10.4.1 What Oximetry Measures Can Tell Us

The pulse oximeter may be used in a variety of situations that require monitoring of oxygen status and may be used either continuously or intermittently. It can give an early warning of decreasing arterial oxyhemoglobin saturation prior to the subject exhibiting clinical signs of hypoxia.

Oximetry may be indicated during exercise testing for evaluation of hypoxemia and/or desaturation in the presence of: a history and physical indicators suggesting hypoxemia and/or desaturation (e.g., dyspnea, pulmonary disease), abnormal diagnostic test results.

4.1.10.4.2 What Oximetry Measures Cannot Tell Us

Pulse oximeters did not work in all situations. Factors that may affect readings, limit precision, or limit the performance or application of a pulse oximeter include motion artifact, abnormal hemoglobins (primarily carboxyhemoglobin [COHb] and met-hemoglobin [metHb]). Other factors are exposure of the measuring probe to ambient light during measurement, low perfusion states and skin pigmentation. Nail polish, varnish or nail coverings when a finger probe is used can reduce the ability to detect saturations below 83% with the same degree of accuracy and precision seen at higher saturations. Anemia (a hemoglobin less than 5) will cause the oximeter to display a false high saturation when the patient is actually hypoxic.

Pulse oximetry cannot distinguish between different forms of haemoglobin. With carbon monoxide poisoning (carboxyhemoglobinemia – haemoglobin combined with carbon monoxide), the pulse oximeter is not able to distinguish oxyhemoglobin from carboxyhemoglobin. Both will be read together and a false high saturation reading will be the result. Carboxyhaemoglobin is registered as 90% oxygenated haemoglobin and 10% desaturated haemoglobin – therefore the oximeter will overestimate the saturation. The presence of methaemoglobin will prevent the oximeter working accurately and the readings will tend towards 85%, regardless of the true saturation. The most important limitation of pulse oximeters is that *they will not detect carbon monoxide (CO) poisoning.* When CO binds with the hemoglobin in blood, the cells turns bright red just as if they had been oxygenated. The resulting molecule (carboxyhememglobin) is incapable of carrying O2 to blood cells, but is indistinguishable in colour from oxygenated blood so far as the pulse oximeter is concerned. Consequently, it's *important for pilots to carry a CO detector*, especially when flying single-engine aircraft that utilize an exhaust-muff-type cabin heat system.

Oximeters give no information about the level of CO_2 and therefore have limitations in the assessment of patients developing respiratory failure due to CO_2 retention.

Methemoglobin is a form of hemoglobin in which the iron has been oxidized and is no longer capable of transporting oxygen. This can occur with exposure to nitrates, nitrites, phenacetin, pyridium, sulfonamides, or benzocaine. Methemoglobin is equally absorbed by both of the oximeter's light wavelengths. This corresponds to a functional saturation of 85% on the curve, which means the reading will tend towards 85% regardless of the true saturation. Therefore, if the functional saturation is really less than 85%, the oximeter will read high, and if the functional saturation is really greater than 85%, the oximeter will read low. Pulse oximeters tend to become inaccurate at extremely low levels of oxygen saturation (below about 75%).

Finally, the pulse oximeter depends on the presence of a good pulse. Subjects with unusually low blood pressure or impaired blood flow to the fingers may have difficulty getting valid oximeter readings. Conditions that cause constriction of the blood vessels in the extremities (e.g., cold temperatures, profound hypoxia) can also interfere with oximeter readings, as can drugs that are vasoconstrictors or vasodilators (e.g., nitroglycerine), or drugs that affect blood color (e.g., sulfonamides). In most such cases, the oximeter will warn of inadequate perfusion (Moller et al., 1993). Although these problems are not often seen in the cockpit.

4.1.10.5 General Advantages/Disadvantages of Oximetry

The great advantages of pulse oximeters are that they are non-invasive and that they make continuous measurements. They measure either the optical absorbance or reflectance of haemoglobin in capillaries in the skin to assess oxygen saturation. Because cutaneous blood flow serves mainly to regulate body temperature, the skin extracts relatively little oxygen from the blood. Therefore, pulse oximeters estimate *the functional saturation of arterial blood.*

Oximeters are calibrated during manufacture and automatically check their internal circuits when they are turned on. They are accurate in the range of oxygen saturations of 70 to 100% (+/-2%), but less accurate under 70%. The pitch of the audible pulse signal falls with reducing values of saturation.

4.1.10.6 Apparatus Required

The past years have seen increased activity by pulse oximeter manufacturers in developing new instruments to improve the reliability of pulse oximetry. Among these are Nellcor NPB-290, NPB-295, and Symphony N-3000 that include Oxismart technology. The new NPB-395 includes Nellcor's new Oxismart XL signal processing and alarm management technologies. The N-395 has been cleared by the FDA to make accuracy claims in adults during motion. No published studies of this new device are currently available.

Oxismart and Oxismart XL are alarm management technologies designed to identify and reject artifacts that could be otherwise mistaken for a pulse. Oxismart XL uses a variety of signal filters intended to reduce error. In the case of Oxismart, when interference is detected, the monitor software continues to search for a pulse as long as continuous artifact is detected. If the pulse oximeter fails to detect at least one qualified pulse in a ten-second period, the display alternates between data and dashes, and a data evaluation period is entered. During the data evaluation period, the last "clean" reading is held in the display until another reading can be attained. This period continues for a total of 60 seconds until the monitor zeroes out and an audible alarm is sounded.

4.1.10.7 Personnel Required

No special personnel are needed to maintain the pulse oximeter readings. Although, the positioning of the sensor probe is important to obtaining an accurate reading, as is using the appropriate type and size of probe. When using a finger probe, utilize the arm not in use for blood pressures, arterial lines, or pressure dressing. The sensor should be placed flush with the skin, and secured without compromising the circulation in the finger. During long term continuous use, the probe site should be checked often to avoid tissue injury. If in doubt of the accuracy of the reading, reposition the sensor or try using another finger. A pulse oximeter permits crewmembers and passengers of an aircraft to evaluate their actual need for supplemental oxygen quickly and easily. There are prepared tutorials and guidelines on application of pulse oximetry in medicine and aviation. As with any such recommendation, each pilot has the obligation to become familiar with the technology and its proper use, and to interpret and adapt these guidelines to the particular situation. Some pilots and passengers will need to use supplemental oxygen at oxygen saturation levels higher than other individuals, and some may need higher oxygen flow rates than others.

4.1.10.8 Application

Pulse oximetry has become the standard technique for monitoring oxygenation during operational mission. This technique has been readily adopted because it easily, continuously and non-invasively provides valuable data on arterial oxygenation (SpO_2) and requires no calibration. Pulse oximetry is widely used as a safety/trending monitor for operators at risk of respiratory distress and/or hypoxemia. During operational missions some operators are at risk for respiratory distress and/or hypoxemia and therefore should be considered candidates for continuous monitoring with pulse oximetry. In some aeromedical circumstances, it is impossible to control cabin altitude.

Pulse oximetry allows to avoid the deadly effects of hypoxia in real-time, by knowing the warning signs of hypoxia and by evaluating the effects of high-altitude on oxygen saturation.

4.1.10.9 References

Ernsting, J. (1984). Mild Hypoxia and the Use of Oxygen in Flight. *Aviation Space Environmental Medicine*, *55*, 407-410.

Moller, J.T., Pedersen, T., Rasmussen, L.S., Jensen, P.F., Pederson, B.D., Ravlo, O. Rassmusen, N.H., & Espersen, K. (1993). Randomized Evaluation of Pulse Oximetry in 20,802 Patients: I. Design, Demography, Pulse Oximetry Failure Rate, and Overall Complication Rate. *Anesthesiology*, *78*, 436-444.

Murrison, A., & Francis, T. (1991). An Introduction to Decompression Illness. *British Journal of Hospital Medicine*, *46*, 107-110.

Nehrenz, G.M. (1997). Altitude and Oxygenation. *The Internet Journal of Aeromedical Transportation*, *1 (1)*, :http://www.ispub.com/journals/IJAMT/Vol1N1/altox.htm.

Pulse Oximetry (1991). AARC Clinical Practice Guideline. *Respiratory Care*, *36*, 1406-1409.

Severinghaus, J.W., & Kelleher, J.F. (1992). Recent Developments in Pulse Oximetry. *Anesthesiology*, *76*, 1018-1038.

Sheffield, P.J., & Heimbach, R.D. (1996). Respiratory Physiology. In R.L. DeHart (Ed.), *Fundamentals of Aerospace Medicine* (pp. 69-108). Baltimore: Williams and Wilkins.

4.1.11 Stress Hormones

4.1.11.1 Background

The response to psychological and physical stress is expressed by changes in hormone levels that can be monitored by assays of body fluids. Two categories of hormones have received the greatest attention, catecholamines and cortisol. Catecholamines were among the first to be studied and have been found to affect peripheral organs. Their levels increase in stressful situations and prolonged increases may have long-term negative effects on health. Cortisol can act directly on the central nervous system because it can cross the blood-brain barrier (Lovallo & Thomas, 2000). The effects of stress hormones are also found in conditions of intense physical workload and when there has been too little recovery time. This, combined with other stressors, can produce signs of fatigue which can result in impairment of mood and performance that can also be observed in burnout and overtraining that is associated with fatigue, a decrease in performance, and lengthening of recovery time, or possibly in conditions of staleness (Eichner, 1995; Hooper, McKinnon, Gordon & Bachmann, 1993). Fatigue can be associated with decreased performance and an impairment of circulatory, immune, metabolic, and hormonal functions. Most of these symptoms result from a disruption of the hypothalamic-pituitary axis (Figure 16).

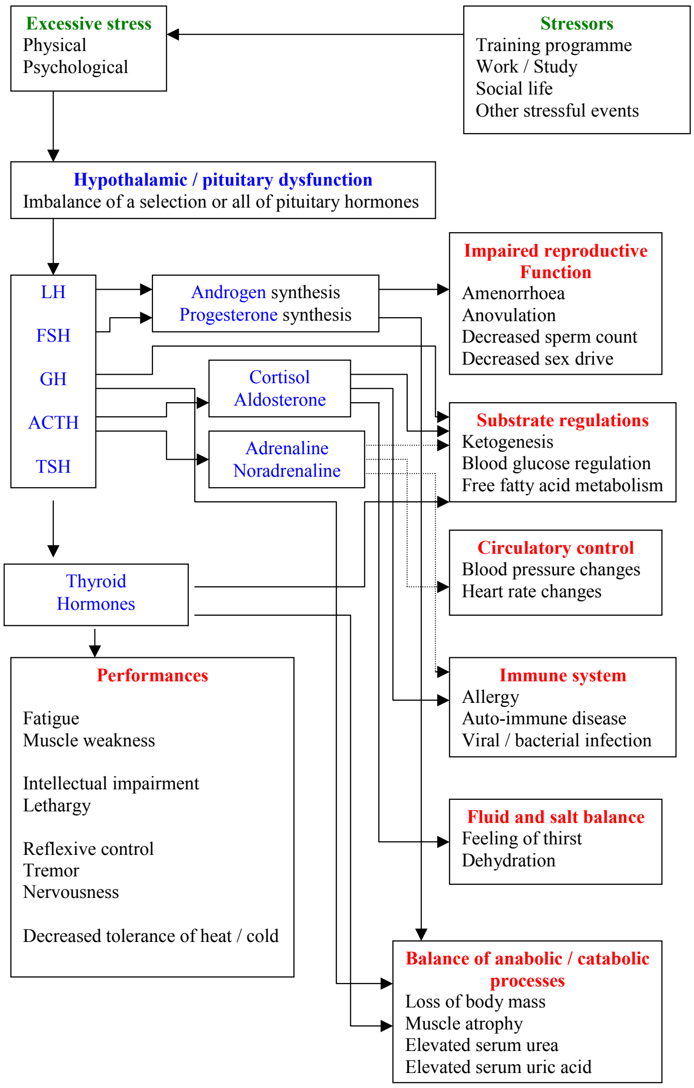

Figure 16: Disruption of the Hypothalamo-Pituitary Axis Produces Most of the Symptoms of Overtraining. (from Fry et al., 1994).

Psychological stress has been shown to result in two different response patterns depending on the person's control over the situation (Kalimo, Lindström & Smith, 1997). For example, effort at the job that produces stress will be accompanied by increased catecholamine secretion but not increased cortisol secretion. Stress and simultaneous distress, such as uncertainty and anxiety, is associated with increases in both catecholamines and cortisol (Frankenhauser & Johansson, 1986).

Plasma and urinary catecholamines are considered to be useful objective markers of stress and fatigue when considered in conjunction with the self-assessment of well-being. Urinary catecholamines measured during resting periods have been reported to show lower levels after overtraining compared to levels prior to the overtraining (Lehmann et al., 1991). Venous norepinephrine may rise after workouts in short periods of resistance exercise (Fry, Kraemer & Van Borselen, 1994) or may decrease with exhaustion in endurance exercise (Lehmann, Gastmann & Petersen, 1992; Lehmann, Schnee, & Scheu, 1992).

Excessive stress leading to imbalances in the neuroendocrine axis may contribute to the overtraining syndrome. Barron and coworkers (Barron, Noakes, Levy, Smith, & Milla, 1985) presented evidence that exhaustion of the hypothalamus, which is less sensitive to the stress of hypoglycemia in overtrained athletes, resulted in impaired responses of ACTH, cortisol, growth hormone (GH), and prolactin to hypoglycemia.

The overtraining syndrome has been associated with high resting blood or salivary levels of cortisol in some studies (Barron et al., 1985; Neary, Wheeler & McLean, 1994; O'Connor, Morgan, Raglin, Barksdale & Kalin, 1989) but not in others where cortisol was found to decrease (Lehmann, Foster, & Keul, 1993; Lehmann, et al., 1992; Lehmann, et al., 1992) or to not change (Flynn, Pizza & Boone, 1994; Hooper, et al., 1993; Verde, Thomas & Shepard, 1992).

Intense physical exercise and overtraining have been shown to decrease free testosterone by direct inhibition of testicular secretion, inhibition of the hypothalamic-pituitary-adrenal-testicular axis, and increase of the sex hormone-binding globulin (SHBG). The latter contributes to the low free testosterone levels by increasing the testosterone binding capacity of serum (Cumming, Wheeler & McColl, 1989). Since testosterone tends to fall in overtrained male athletes, the testosterone/cortisol ratio has also been thought to be a marker of fatigue. But this ratio in overtraining may also decrease (Vervoorn, Quist & Vermulst, 1991) or may be steady (Flynn et al., 1994) and is of little use in females. In fact, the testosterone/cortisol ratio is related to intense and prolonged physical or mental exercise rather than to fatigue per se (Karvonen, 1992), so it is no longer considered a reliable marker of fatigue.

Further support for hypothalamic-pituitary dysfunction in overtrained or fatigued people is provided through female athletes who have been shown to develop amenorrhoea through exercise (Loucks, 1990). Impaired ovarian function was thought to be related to decreased pituitary hormone secretion (Feicht, Johnson, & Martin, 1978) due to alteration in the hypothalamic control of gonadotrophin release (GHRH) (McArthur, Bullen & Beitens, 1980). Other markers of overtraining may include decrease in the pulsatile nature of LH and FSH release, luteal phase shortening, changes in gonadal steroid concentrations (estradiol, progesterone, or testosterone), low T3 levels, lack of appropriate thyroid-stimulating hormone (TSH) response to thyrotropin-releasing hormone (TRH), and increase in endorphine secretion (also resulting in a decrease in the FSH and LH secretions; Cumming, Vickovic, Wall, & Flucker, 1985).

Additionally, performance is influenced by an endogenous circadian component that drives the same pacemaker controlling other physiological rhythms, including plasma cortisol and melatonin (Monk et al., 1997). Within subjects, predominantly negative correlations were found between good performance and higher plasma levels of cortisol and melatonin. Thus, these hormones, which are not reliable and direct markers of fatigue, are strong indicators of the circadian component of performance and vigilance.

4.1.11.2 State of the Art

Plasma cortisol and melatonin have already been measured from salivary samples during a real-world study on jet lag (called "Pegasus operation"). It was an experimentation conducted in collaboration between American and French scientists to investigate the potential caffeine-induced resynchronization of endogenous melatonin and cortisol secretions (Piérard et al., 2001).

4.1.11.3 Measurement of Melatonin

Melatonin can be measured form saliva samples that are collected in darkness upon awaking using Sarstedt® "salivettes". A cotton pad is put under the tongue for a few minutes until completely saturated. It is then put in the upper compartment of a two-compartment polyethylene tube. The two compartments communicate by a hole, allowing the collection of saliva in the lower compartment when the tube is centrifuged. The samples are centrifuged and then transferred to 1 ml cryotubes and frozen at -20°C.

Salivary melatonin is quantified by the means of radioimmunoassay (RIA). In order to avoid contamination, precautions are taken with regard to glassware cleanness and distillated water quality.

4.1.11.4 Measurement of Cortisol

Free salivary cortisol, as a surrogate of plasma cortisol levels, can also be quantified from saliva as for melatonin on waking. Maximal cortisol secretion occurs between 6 a.m. and 9 a.m., thereafter decreasing continuously until midnight, thus providing a strong chronobiologic marker. However, darkness is not necessary when sampling. Processing and storage procedures are the same as for melatonin.

Cortisol can be assayed by solid-phase radioimmunoassay (RIA). Fluorimetry can also be used to measure plasma cortisol. This classical, sensitive (100 nmol/l of cortisol), simple, and fast (5 min) method uses one ml of blood. The RIA can be processed in the field using a small fluorimeter, but it does require pure reagents.

Free cortisol can also be measured from urine using gas chromatography. This can be easily accomplished in the field using portable chromatographs that need to be connected to gas containers or generators.

4.1.11.5 Measurement of Testosterone, Estrogens and Progesterone

Free plasma testosterone, E2 estradiol, and progesterone levels are measured only by RIA using specific antibodies (see above). The secretion of estradiol and progesterone can also be assessed by measuring their metabolites in urine. The most accurate method is the High Performance Liquid Chromatography method. However, analysis can also be done using colorimetric methods. For example, the progesterone's metabolites (pregnanediol and pregnanetriol) give a yellow color when they are hydrolyzed in a sulfuric solution (Talbot's reaction).

4.1.11.6 Measurement of the Plasma Hypothalamic Stimulating Hormones

Plasma levels of Growth Hormone (GH), Prolactin (PRL), Follicle Stimulating Hormone (FSH), Luteinizing Hormone (LH), and Thyroid Stimulating Hormone (TSH) are measured using the RIA or the Enzyme Linked ImmunoSorbent Assay (ELISA) methods. Both methods need a specific antibody that is radiolabeled (^{125}I) for RIA or linked to an enzyme (peroxydase or alcaline phosphatase) for ELISA.

4.1.11.7 Possibilities and Limitations

4.1.11.7.1 Possibilities

Fatigue, staleness, and overtraining syndrome have all been considered as a generalized stress response and as a neuroendocrine disorder. Therefore, stress hormones have been monitored in an attempt to understand possible mechanisms and to find reliable indicators of the disorder. In this context it would be useful to assay norepinephrine plasma level and urinary excretion, which are expected to respectively increase (at least in resistance-exercise) and decrease during fatigue. Cortisol is not a good indicator of fatigue, but on the other hand, it is established that this hormone can be considered, like melatonin, as a good marker of circadian rhythms. Since any shift of these rhythms results in a fatigue syndrome, plasma cortisol and melatonin are also of interest to assess operator functional state.

4.1.11.7.2 Limitations

In fact, fatigue and its likely influence on performance cannot be directly assessed using hormonal data exclusively. For example, central fatigue must be estimated using tests exhibiting numerous symptoms, such as disturbance of perception, coordination, activity, motivation, and performance. Moreover, in the absence of fatigue, plasma catecholamines are correlated with psychological stress, which also depends on the personality.

4.1.11.8 General Advantages/Disadvantages of Hormonal Analysis

Hormonal data obtained from salivary samples have known correlations with blood samples, so that the cortisol and melatonin analysis from saliva represents an easy, quick, reliable, and accurate method of circadian rhythm assessment.

Measures of catecholamines could be possible in baseline and even current conditions of the OFS.

Because of the large range of inter-individual and intra-individual variability of hormonal data, hormonal analysis must be processed under standardized conditions. Saliva samples are preferred, because venopuncture is less acceptable to operators and the blood samples are heavier. Blood samples are needed for catecholamines. Samples must be obtained just after waking with lights off for melatonin, or early in the morning for cortisol and catecholamines.

4.1.11.9 Apparatus Required

Before analysis, samples must be centrifuged. The most accurate measurements of hypothalamic (GH, FSH, LH, TRH, PRL) and gonadal hormones (testosterone, progesterone, estradiol) are obtained using the RIA and HPLC methods. The most up-to-date analyzers can be set up in a shelter, thus allowing on-line analysis in the field. If this equipment is not available, plasma, urinary and saliva samples must be frozen quickly to at least -20°C; which must be maintained using dry ice or liquid nitrogen during shipping to the laboratory.

Measures of cortisol and catecholamine levels can also be processed in the field using a portable fluorimeter or chromatograph, but this requires pure reagents and very clean glassware. In contrast, for melatonin analysis, saliva samples must be shipped to a laboratory.

4.1.11.10 Personnel Required

Trained laboratory personnel are needed for collecting samples, preparing the samples for analysis, and for the analysis.

4.1.11.11 References

Barron, J.L., Noakes, T.D., Levy, W., Smith, C., & Millar, R.P. (1985). Hypothalamic Dysfunction in Overtrained Athletes. *Journal of Clinical Endocrinology and Metabolism*, *60*, 803-806.

Cumming, D.C., Vickovic, M.M., Wall, S.R., & Flucker, M.R. (1985). Defects in Pulsatile LH Release in Normally Menstruated Runners. *Journal of Clinical Endocrinology and Metabolism*, *60*, 810-812.

Cumming, D.C., Wheeler, G.D., & McColl, E.M. (1989). The Effects of Exercise on Reproductive Function in Men. *Sports Medicine*, *7*, 1-17.

Eichner, E.R. (1995). Overtraining: Consequences and Prevention. *Journal of Sports Sciences*, *13*, S41-S48.

Feicht, C.B., Johnson, T.S., & Martin, B.S. (1978). Secondary Amenorrhoea in Athletes. *Lancet*, *2*, 1145-1146.

Flynn, M.G., Pizza, F.X., & Boone, J.B. (1994). Indices of Training Stress during Competitive Running and Swimming Seasons. *International Journal of Sports Medicine*, *15*, 21-26.

Fry, A.C., Kraemer, W.J., & Van Borselen, F. (1994). Catecholamine Responses to Short-Term High-Intensity Resistance Exercise Overtraining. *Journal of Applied Physiology*, *77*, 941-946.

Frankenhauser, M., & Johansson, G. (1986). Stress at Work: Psychobiological and Psychosocial Aspects. *International Review of Applied Psycholo*gy, *35*, 287-299.

Hooper, S.L., McKinnon, L.T., Gordon, R., & Bachmann, A.W. (1993). Hormonal Responses of Elite Swimmers to Overtraining. *Medicine and Science in Sports and Exercise*, *25*, 741-747.

Kalimo, R., Lindström, K., & Smith, M.J. (1997). Psychosocial Approach in Occupational Health. In: G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1058-1084). New York: John Wiley & Sons.

Karvonen, J. (1992). Overtraining. In J. Karvonen, P.W.R. Lemon, & I. Iliev (Eds.), *Medicine in Sports Training and Coaching* (Vol. 35, pp. 174-188). Basel: Karger.

Lehmann, M., Dickruth, H.H., Gendrisch, G., Lazar, W., Thum, M., Kaminski, R., Aramendi, J.F., Peterke, E., Wieland, W., & Keul, J. (1991). Training-Overtraining: A Prospective, Experimental Study with Experienced Middle- and Long-Distance Runners. *International Journal of Sports Medicine*, *12*, 444-452.

Lehmann, M., Foster, C., & Keul, J. (1993). Overtraining in Endurance Athletes: A Brief Review. *Medicine and Science in Sports and Exercise*, *25*, 854-862.

Lehmann, M., Gastmann, U., & Petersen, K.G. (1992). Training-Overtraining: Performance and Hormone Levels after a Defined Increase in Training Volume vs. Intensity in Experienced Middle- and Long-Distance Runners. *British Journal of Sports and Medicine*, *26*, 233-242.

Lehmann, M., Schnee, W., & Scheu, R. (1992). Decreased Nocturnal Catecholamine Excretion: Parameter for Overtraining Syndrome in Athletes? *International Journal of Sports Medicine*, *13*, 236-242.

Lovallo, W.R., & Thomas, T.L. (2000). Stress Hormones in Psychophysiological Research: Emotional, Behavioral, and Cognitive Implications. In J.T. Cacioppo, L.G. Tassinary & G.G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 342-367). Cambridge: Cambridge University Press.

Loucks, A.B. (1990). Effect of Exercise Training on the Menstrual Cycle: Existence and Mechanisms. *Medicine and Science in Sports and Exercise*, *22*, 275-280.

McArthur, J.W., Bullen, B.A., & Beitens, I.Z. (1980). Hypothalamic Amenorrhoea in Runners of Normal Composition. *Endocrinological Research Communications*, *7*, 13-25.

Monk, T., Buysse, D.J., Reynolds, C.M., Berga, S.L., Jarrett, D.B., Begley, A.E., & Kupfer, D.J. (1997). Circadian Rhythms in Human Performance and Mood under Constant Conditions. *Journal of Sleep Research*, *6*, 9-18.

Neary, J.P., Wheeler, G.D., & McLean, I. (1994). Urinary Free Cortisol as an Indicator of Exercise Training Stress. *Clinical Journal of Sports Medicine*, *4*, 160-165.

O'Connor, P.J., Morgan, W.P., Raglin, J.S., Barksdale, C.M., & Kalin, N.H. (1989). Mood State and Salivary Cortisol Levels following Overtraining in Female Swimmers. *Psychoneuroendocrinology*, *14*, 303-310.

Piérard, C., Beaumont, M., Enslen, M., Chauffard, F., Tan, D.X., Reiter, R.J., Fontan, A., French, J., Coste, O., & Lagarde, D. (2001). Resynchronization of Hormonal Rhythms after an Eastbound Flight in Humans: Effects of Slow-Release Caffeine and Melatonin. *European Journal of Applied Physiology*, *85*, 144-150.

Verde, T., Thomas, S., & Shepard, R.J. (1992). Potential Markers of Heavy Training in Highly Trained Distance Runners. *British Journal of Sports and Medicine*, *26*, 167-175.

Vervoorn, C., Quist, A.M., & Vermulst, L.J.M. (1991). The Behaviour of the Plasma Free Testosterone/Cortisol Ratio during a Season of Elite Rowing Training. *International Journal of Sports Medicine*, *12*, 257-263.

4.2 PERFORMANCE TESTS

4.2.1 Background

Performance tests are used in numerous applications including (1) personnel selection (to predict work performance or eventual job/career success), (2) training (to assess training effectiveness and predict future performance), (3) risk factor evaluation (to determine the effects of stressors such as alcohol, antihistamines and other chemical agents, sustained operations, and harsh environments), and (4) the assessment of readiness to perform or the more general OFS. Many computer-based tests are historically founded in traditional pencil-and-paper tests of cognitive ability, several of which are still in use. Other tests take advantage of unique capabilities afforded by a computer-based test, such as millisecond response timing, dynamic movement for tracking and monitoring tasks, and the simultaneous presentation of multiple tasks to assess attention and time-sharing resources.

4.2.2 Rationale

A number of fundamental assumptions must be made concerning the use of performance tests to measure operator functional state:

1. By measuring processes common to both tasks, performance on an assessment task is indicative of performance on a work/job task.
2. At the very least, *changes* in performance on the assessment task are indicative of *changes* in performance on the work/job task.
3. Because of regulatory control, *performance may be protected* by recruitment of effort, so that decrements may not be revealed in primary tasks.
4. The *costs* associated with performance protection may be taken as indicative of *latent decrement* (i.e., that primary task performance demands greater effort and attention, and is more vulnerable to disruption from further load/stress).

The validity of the assumptions relies on a number of factors including comprehensiveness of the assessment task (context validity), adequate task training (to attain asymptotic levels of performance), and high levels of engagement (effort and motivation – the ability of the task to engage the participant, as in real-life tasks). Measurement of performance must be accompanied by measurement of cost/effort/psychophysiological state in order to detect subtle changes (e.g., a shift in the performance/cost trade-off function) and *latent* decrement (i.e., a reduction in efficiency resulting from an increase in costs for measured performance). Alternatively, where cost/effort is already maximal (e.g., in well-trained operational contexts), detection of a decrement may require the use of tasks that demand maximal effort, or secondary tasks that assess workload and spare capacity.

4.2.3 State of the Art

A comprehensive review of existing computer-based tasks and performance assessment batteries would be very lengthy. The current review focuses on categories of tests and provides a representative sample in each category. For recently developed and released tests, normative data may not yet be available. Although now a decade old, an excellent review of computer-based tests used for neuropsychological and performance-based assessment was provided by Kane and Kay (1992). In their review, thirteen major computer-based cognitive performance assessment batteries were examined with information provided on (1) development history, (2) hardware requirements, (3) included tasks, (4) test administration, (5) parameter options, (6) data output, (7) norms, and (8) validation studies. These topics provide valuable information on test limitations. Information is also provided on individual tests common to several batteries. The following taxonomy was used to classify the individual tests: Simple Motor Tests, Reaction Time Tests, Attention-Concentration Working Memory, Learning and Memory, Spatial Perception/Reasoning, Calculations, Language, Complex Problem Solving, Dual-Tasking and Multi-Tasking. Other researchers have provided similar reviews (Horst & Kay, 1988, 1991; Kay & Horst, 1988).

4.2.3.1 Simple Cognitive/Psychomotor Tests

Historically, the Performance Evaluation Tests for Environmental Research (PETER, U.S. Navy; Bittner, Carter, Kennedy, Harbeson, & Krause, 1986), the Bexley-Maudsley Automated Psychological Screening (B-MAPS; Acker & Acker, 1982), the Criterion Task Set (CTS, U.S. Air Force; Shingledecker, 1984; Schlegel & Gilliland, 1990), and the Walter Reed Performance Assessment Battery (WRPAB, U.S. Army; Thorne, Genser, Sing, & Hegge, 1985) were among the first collections of simple performance assessment tests developed. Originally programmed on Apple II and Commodore level machines, the included tests addressed such classic cognitive psychology abilities as display monitoring, memory recall and recognition using various symbol domains, grammatical reasoning, spatial processing, pattern comparison, category sorting, mathematical processing, linguistic processing, and visuomotor tracking.

The tests in these foundational batteries, along with other tests of simple cognitive processing, have been incorporated in numerous collections, including the Automated Portable Test System (APTS; Bittner, Smith, Kennedy, Staley, & Harbeson, 1985; Kennedy, Dunlap, & Kuntz, 1989), the Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB; Englund et al., 1987; Hegge, Reeves, Poole, & Thorne, 1985; Schlegel & Gilliland, 1992), the NATO AGARD-STRES (Standardized Tests for Research with Environmental Stressors; Santucci et al., 1989; Reeves et al., 1991), which contained components from the CTS, UTC-PAB and the TNO TaskOMat, the Automated Neuropsychological Assessment Metrics (ANAM; Reeves et al., 1992) for clinical neurological screening, a subset of ANAM configured as the Spaceflight Cognitive Assessment Tool (WinSCAT) used by NASA for assessing severe neurological effects of space flight incidents, and COGSCREEN used by the Federal Aviation Administration for detecting changes in the cognitive functioning of aviators (Horst & Kay, 1991). Recent commercial additions for concussion evaluation include Headminders and Impact.

The general approach in simple test configuration is to present to the participant a series of stimuli, each requiring processing according to the rules of the test, and to measure response latency and accuracy. As such, the tests are usually effective for identifying the impact of significant stressor levels (especially when evaluating group effects), but lack sensitivity at lower stressor levels and when evaluating the functional state of individuals. More details about the various task batteries can be found in Gilliland and Schlegel (1993).

4.2.3.2 Complex Tasks: Time-Sharing and Divided Attention

Although originally developed using mechanical components and hardwired logic, the Multiple Task Performance Battery (MTPB) represents an early implementation of a sophisticated multiple-task performance assessment tool. The MTPB provided assessment of monitoring, arithmetic, and complex code-solving performance in a time-sharing work environment (Chiles & Jennings, 1970; Chiles, Alluisi, & Adams, 1968; Chiles, Iampietro, & Higgins, 1972; Chiles, Jennings, & West, 1972). It also served as a model for later multiple-task tests such as the Synthetic Work Task (SYNWORK), the NASA Multi-Attribute Task Battery (MATB; Arnegard, 1990, 1991), and component tasks in the NASA Performance Assessment Workstation (PAWS). Dual tasks involving visuomotor tracking and memory have been included in a number of popular test batteries due to their ability to tap both cognitive and psychomotor processes simultaneously. Because of the additional demands placed on resource allocation and scheduling, these tasks often provide the sensitivity needed to identify smaller changes in operator functional state.

A commercial product called NovaScan provides a directed attention test in which a degraded state is indicated if performance declines relative to baseline when an individual shifts resources from one task type or process to another. The elemental tasks may be changed to be relevant to components of the work task.

4.2.3.3 Complex Tasks: Work Samples and Simulations

The NASA Multi-Attribute Task Battery (MATB) was developed to provide a comprehensive behavioral metric for assessing operator performance, and was structured to approximate an aircrew operations environment (Arnegard, 1990, 1991). Like the FAA Air Traffic Scenarios Test (ATST) used to screen air traffic controller candidates, the MATB provides an integrated setting of component tasks that offer high levels of engagement due to their similarity to real jobs. A more distinctive and engaging set of simulations are those referred to as ‘microworlds’ (e.g., Brehmer, Leplat, & Rasmussen, 1991; Dorner, 1987) – high fidelity computer simulations of complex work environments such as firefighting, urban planning, or process control, as much concerned with the analysis of strategy and tactics (i.e., *how* humans solve operational problems) as in effectiveness per se. Hockey’s (1997) CAMS (Cabin Air Management System) assesses both effectiveness and strategic behavior. Hockey, Wastell, and Sauer (1998) showed graded effects on secondary tasks and costs of sleep deprivation and interface dialogue control, in the absence of effects of overt performance, and a clear relation to effort involvement of operators. CAMS has also been used to detect the effects of fatigue in extended periods of exposure to extreme environments such as Antarctic over-wintering and simulated space flight (Sauer, Hockey, & Wastell, 1999a, b). Parasuraman and colleagues (Lorenz, Di Nocera, & Parasuraman, 2001) have recently adapted CAMS to allow the level of human vs. machine dialogue control to be manipulated, and are currently using it to study trade-offs in adaptive automation.

4.2.4 Possibilities and Limitations

Performance tests cannot give an absolute index of mental competence, or even of motor or perceptual skill. Performance is not the same thing as efficiency, although this equivalence is assumed in many research papers, even in good journals. In its usual form (task batteries, simple cognitive tests, etc.)

performance can be seen as an index of *effectiveness* (i.e., how well specific task goals are being met). It can tell us something about the *efficiency* of the mental processes only if we measure the costs of maintaining that level of effectiveness. With these caveats, performance tests can be designed to provide powerful, valid, and reliable indices of changes in both manifest and latent degradation. The appropriate use of auxiliary measures reflecting costs (e.g., effort, post-task fatigue, and physiological activation of various kinds) will permit more strongly diagnostic inferences to be made about the impact of a task or environmental variable on the operator.

4.2.4.1 Limitations

Adequate training – Assessment of performance changes within persons (the only interesting kind?) is often hampered by the presence of large learning effects during repeated testing because of low levels of initial training. It is necessary to train on all tasks so that learning is near asymptotic. Then, changes with factors that affect either low-level processes or control mechanisms should give rise to detectable increases or decreases around a steady state of performance.

Motivation – This is partly connected to the poor training issue, but also the need to ensure task involvement (as in real-life, work, etc.). Overt performance decrements in most stress studies are the result of reduced motivation in tasks that do not matter much to the participants – they just stop making the effort when the going gets tough. Of course this is an important effect, but is an effect on task engagement (maintaining the task goal in the driving seat), rather than on cognitive processes per se. Motivation is also highly variable and produces high variance data. If motivation is kept high (as in operational tasks), one can measure the strain of performance protection more effectively as the spillover into costs.

Measurement context – Care must be taken to recognize performance assessment for what it is – just one component of the overall multivariate response to task (and environmental) demands. A system-oriented approach takes into account the impact of situational variables on both performance and other components – physiology, subjective state, and impact on technical system resources (e.g., more wasteful use of energy/power, accidents, and stoppages) – all are part of the overall system adaptive response. In the extreme case, task performance may seem fine, while the operator feels frantic, his or her physiology is through the roof, power use is profligate and the system keeps crashing.

4.2.5 References

Acker, W., & Acker, C. (1982). Bexley Maudsley Automated Psychological Screening and Bexley Maudsley Category Sorting Test: Manual. Windsor, Great Britain: NFER-Nelson.

Arnegard, R.J. (1990). *Multi-Attribute Task Battery: The Bottom Line*. Poster presented at the 34th Annual Meeting of the Human Factors Society, Orlando, FL.

Arnegard, R.J. (1991). *Operator Strategies under Varying Conditions of Workload*. National Aeronautics and Space Administration, NASA Contractor Report 4385, NASA Langley Research Center.

Bittner, A.C., Jr., Carter, R.C., Kennedy, R.S., Harbeson, M.M., & Krause, M. (1986). Performance Evaluation Tests for Environmental Research (PETER): Evaluation of 114 measures. *Perceptual and Motor Skills, 63,* 683-708.

Bittner, A.C., Jr., Smith, M.G., Kennedy, R.S., Staley, C.F., & Harbeson, M.M. (1985). Automated Portable Test System (APTS): Overview and Prospectus. *Behavioral Research Methods, Instruments, and Computers, 17,* 217-221.

Brehmer, B., Leplat, J., & Rasmussen, J. (1991). Use of Simulation in the Study of Complex Decision Making, In J. Rasmussen, B. Brehmer, & J. Leplat (Eds.), *Distributed Decision Making: Cognitive Models for Cooperative Work* (pp. 373-386). New York: John Wiley & Sons.

Chiles, W.D., & Jennings, A.E. (1970). Effects of Alcohol on Complex Performance. *Human Factors, 12,* 605-612.

Chiles, W.D., Alluisi, E.A., & Adams, O.S. (1968). Work Schedules and Performance during Confinement. *Human Factors, 10,* 143-196.

Chiles, W.D., Iampietro, P.F., & Higgins, E.A. (1972). Combined Effects of Altitude and High Temperature on Complex Performance. *Human Factors, 14,* 161-172.

Chiles, W.D., Jennings, A.E., & West, G. (1972). *Multiple Task Performance as a Predictor of the Potential of Air Traffic Controller Trainees* (Report AM-72-5). Federal Aviation Administration Office of Aviation Medicine.

Dorner, D. (1987). On the Difficulties People have in Dealing with Complexity. In J. Rasmussen et al. (Eds.), *New Technology and Human Error*. Chichester: Wiley.

Englund, C.E., Reeves, D.L., Shingledecker, C.A., Thorne, D.R., Wilson, K.P., & Hegge, F.W. (1987). *Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB) I: Design and Specification of the Battery.* San Diego, CA: Naval Health Research Center.

Gilliland, K., & Schlegel, R.E. (1993). *Readiness-to-Perform Testing: A Critical Analysis of the Concept and Current Practices* Washington, DC: Department of Transportation, Federal Aviation Administration; 1993; FAA Report No. FAA-AM-93-13. Available from: National Technical Information Service, Springfield, VA 22161. Order # ADA269397.

Hegge, F.W., Reeves, D.L., Poole, D.P., & Thorne, D.R. (1985). *The Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB) II: Hardware/Software Design and Specification*, Fort Detrick, MD: U.S. Army Medical Research and Development Command.

Hockey, G.R.J. (1986). Changes in Operator Efficiency as a Function of Environmental Stress, Fatigue and Circadian Rhythms. In K. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of Perception and Human Performance, Vol. 2: Cognitive Processes and Performance.* New York: Wiley, 44:1-49.

Hockey, G.R.J. (1997). Compensatory Control in the Regulation of Human Performance under Stress and High Workload: A Cognitive Energetical Framework. *Biological Psychology, 45,* 73-93.

Hockey, G.R.J., Wastell, D.G., & Sauer, J. (1998). Effects of Sleep Deprivation and User-Interface on Complex Performance: A Multilevel Analysis of Compensatory Control. *Human Factors, 40,* 233-253.

Horst, R.L., & Kay, G.G. (1988). *Report of Comparative Studies of Cognitive Tests: Cognitive Function Evaluation in Medical Certification of Airmen* (Report FAA/730-42b). Oklahoma City, OK: Federal Aviation Administration.

Horst, R.L., & Kay, G.G. (1991). *Cognitive Function Evaluation in Medical Certification of Airmen: Development and Validation of a Prototype Test Battery* (Report FAA/933-014-90). Oklahoma City, OK: Federal Aviation Administration.

Kane, R.L., & Kay, G.G. (1992). Computerized Assessment in Neuropsychology: A Review of Tests and Test Batteries. *Neuropsychology Review, 3,* 1-118.

Kay, G.G., & Horst, R.L. (1988). Evaluating Cognitive Function: A Review of Mental Status Tests, Neuropsychological Procedures, and Performance Based Approaches (Report on Contract DTFA-02-87-C-87069). Oklahoma City, OK: Federal Aviation Administration.

Kennedy, R.S., Dunlap, W.P., & Kuntz, L.A. (1989). Application of a Portable Automated Performance Test Battery for the Study of Drugs and Driving Performance. Paper presented at the Second International Symposium on Medicinal Drugs and Driving Performance. NOT CITED

Lorenz, B., Di Nocera, F., & Parasuraman, R. (2001). Human Performance during Simulated Space Operations under Varied Levels of System Autonomy. Abstract #020. Proceedings of the Bioastronautics Investigators' Workshop, January 17-19, 2002, Galveston, TX (CD-ROM).

Reeves, D.L., Winter, K., LaCour, S., Raynsford, K., Kay, G., Elsmore, T., & Hegge, F.W. (1992). Automated Neuropsychological Assessment Metrics Documentation: Vol. I Test Administration Guide. Silver Spring, MD: Office of Military Performance Assessment Technology.

Reeves, D.L., Winter, K.P., LaCour, S.J., Raynsford, K.M., Vogel, K., & Grissett, J.D. (1991). The UTC-PAB/AGARD-STRES Battery: User's Manual and System Documentation (Tech. Report NAMRL-SR91-1). Washington, DC: Office of Military Performance Assessment Technology, Walter Reed Army Institute of Research.

Santucci, G., Farmer, E., Grissett, J., Wetherell, A., Boer, L., Gotters, K., Schwartz, E., & Wilson, G. (1989). AGARDograph #308, Human Performance Assessment Methods (ISBN 92-835-0510-7). Neuilly-sur-Seine, France: North Atlantic Treaty Organisation Advisory Group for Aerospace Research and Development, Working Group 12.

Sauer, J., Hockey, G.R.J., & Wastell, D. (1999a). Maintenance of Complex Performance during a 135-day Spaceflight Simulation. *Aviation, Space, and Environmental Medicine, 70,* 236-244.

Sauer, J., Hockey, G.R.J., & Wastell, D. (1999b). Performance Evaluation in Analogue Environments: Adaptation during an 8-month Antarctic Over-Wintering Expedition. *Aviation, Space, and Environmental Medicine, 70,* 230-235.

Schlegel, R.E., & Gilliland, K. (1990). Evaluation of the Criterion Task Set – Part I: CTS Performance and SWAT Data – Baseline Conditions (AAMRL-TR-90-007). Wright-Patterson AFB, OH: Armstrong Aerospace Medical Research Laboratory.

Schlegel, R.E., & Gilliland, K. (1992). Development of the UTC-PAB Normative Database (Tech. Report AL-TR-92). Wright-Patterson AFB, OH: Armstrong Laboratory.

Shingledecker, C.A. (1984). A Task Battery for Applied Human Performance Assessment Research (Tech. Report AFAMRL-TR-84-071). Wright-Patterson AFB, OH: Air Force Aerospace Medical Research Laboratory.

Thorne, D., Genser, S., Sing, H., & Hegge, F. (1985). The Walter Reed Performance Assessment Battery. *Neurobehavioral Toxicology and Teratology, 7,* 415-418.

4.3 SUBJECTIVE MEASURES

4.3.1 Background

Subjective methods are used in human engineering to assess an operator's or observer's task-related experiences with respect to situational awareness (SA), mental workload, and mood states. In principle, either the operator or an observer makes a subjective assessment of the operator's experience expressed and assigns this a numerical value. This assessment is performed by providing ratings on one or more dichotomous scale(s) that are defined by factors describing the possible range of individual experiences. Other techniques that are used to evaluate subjective opinions (e.g., structured interviews) are not treated here because they provide "only" qualitative statements about the characteristics covered by the questionnaire. Therefore, the results cannot be further analyzed by means of statistical techniques.

The widespread use of subjective techniques, especially in mental workload assessment, can be explained because they are easy to implement, non-intrusive, inexpensive, have high face validity, and have proven sensitivity to various demand manipulations in complex systems (O'Donnell & Eggemeier, 1986; Wierwille, & Eggemeier, 1993). The theoretical basis for the sensitivity of subjective measures to changes in workload is the assumption that subjective feelings of task demands, effort, exertion, tension, anger, depression, and confusion can be reported accurately by the subject, and are valid and sensitive indicators of mental workload. Johannsen et al. (1979).

Techniques for the subjective evaluation of effort, exertion, or mood by means of ratings can be subdivided according to several dimensions. One scheme involves the scale type used, which can be nominal, ordinal, interval, or ratio (Lysaght et al., 1989). The type of scale influences the selection of statistical data analysis procedures (parametric vs. non-parametric methods). Interval- and ratio-scaled data can be analyzed by means of parametric tests that make the most use of the data content. However, most subjective techniques provide either ordinal- or interval-scaled data.

The scale dimensionality may be unidimensional, multidimensional, or hierarchical (Hart & Wickens, 1990). Both unidimensional and multidimensional techniques use bipolar scales to obtain individual scores on one or more dimensions of workload. The dimensionality influences the diagnosticity of the technique. Unidimensional techniques provide a global workload value but no information about the source of mental workload. Using multidimensional scaling techniques, several aspects of mental workload are rated separately on ordinal scales. By means of conjoint measurement, a combined metric with interval properties can be generated. The multidimensional techniques allow an identification of the source of workload (i.e., diagnosticity). However the complexity and time required to complete the rating procedure increases with the dimensionality (i.e., the number of scales). Hierarchical scales are handled stepwise and also do not provide information related to workload sources. Additionally, subjective techniques frequently make use of psychometric methods such as magnitude estimation, paired comparisons, and equal-appearing intervals (Pfendler et al., 1995).

Many expert users of workload assessment measures have stated that several considerations should be taken into account in employing subjective techniques. These considerations are of course also valid for the subjective assessment of other issues like situational awareness or mood state. Subjective statements about workload are influenced by factors specific to the task or the operator. Unfortunately, there is no extensive database available that describes the factors influencing subjective workload experience and assessment. For this reason it is often difficult to compare results across studies if these uncontrolled variables have a marked influence. The close connection between mental capacity utilization on the one hand and subjective effort on the other has not been verified, so that subjective techniques are not generally validated. Additionally, a number of individual rating techniques have been developed, which leads to a variety of non-standardized techniques with somewhat limited validation.

4.3.2 State of the Art

There are a wide variety of subjective techniques available in different languages and used for the assessment of mental workload, situational awareness, and mood states. These were generally developed as paper-and-pencil tests, but computerized versions have been implemented for most. Only examples of techniques available in the English language with references for further study are mentioned here.

For some types of subjective measure, rationale and further details are provided in other sections: mental workload (Cognitive Load); situational awareness; sleepiness Sleep Loss). A representative list of tests and methods for these is presented below.

4.3.2.1 Mental Workload

NASA Task Load Index (NASA-TLX)	NASA (1986) Hart & Staveland (1988)
Modified Cooper-Harper Scale (MCH)	Wierwille & Casali (1983)
Sequential Judgment Scale (ZEIS)	Pitrella & Käppler (1988)
Subjective Workload Assessment Technique (SWAT)	Armstrong Aerospace Medical Research Laboratory (1987) Reid et al. (1981) Reid et al. (1989)
Subjective Workload Dominance Technique (SWORD)	Vidulich (1989)

4.3.2.2 Situational Awareness

Situational Awareness Rating Technique (SART)	Taylor (1990)
Crew Awareness Rating Scale (CARS)	McGuinness (1999)
Situation Awareness Global Assessment Technique (SAGAT)	Endsley (1995)
Quantitative Analysis of Situational Awareness (QUASA)	Edgar (2000)

4.3.2.3 Mood States and Task Engagement

Rationale – dimensionality of mood, patterns of response to stressors and control, models of task engagement – demaned//anxiety and fatigue as strain dimensions; effort as nmoderator of workload-fatigue relationship

Profile of Mood States (POMS)	McNair et al. (1971)
PANAS (Positive and Negative Affect Schedule)	Watson, Clark & Tellegen (1988)
UWIST Mood Adjective Check List	Mattews et al. (1990)

4.3.2.4 Limitations

4.3.2.4.1 What Subjective Measures Can Tell Us

Subjective techniques have been shown to be sensitive and valid measures of mental workload, situation awareness, and mood state. Additionally, multi-dimensional techniques may provide diagnostic information with respect to the reason for changing levels of workload. However, in experiments, the ratings along different dimensions have shown high intercorrelations for some techniques, so that diagnostic conclusions have to be drawn carefully. Some techniques can be applied prognostically during system development.

4.3.2.4.2 What Subjective Measures Cannot Tell Us

Unidimensional and hierarchical techniques do not provide any diagnosticity. Subjective measures, as the name implies, can be affected by factors that are not related to aspects of workload (e.g., rating tendencies or bias, response sets, errors, or pre-test attitudes). Additionally, it is possible that subjects intentionally manipulate rating values.

Ratings can only be applied periodically. They do not provide continuous data. Since rating results can also be affected by task-related factors, comparisons of tasks should not be made on the basis of subjective rating results if the task conditions differ to a great extent.

4.3.2.5 General Advantages/Disadvantages

Subjective techniques usually show high levels of inter-individual variability. In order to minimize these effects, training procedures should be applied carefully. Training procedures are usually simple and fast. Only a few techniques require extensive preparation for participants (e.g., SWAT).

Data acquisition during task performance may be intrusive to the primary task, especially with some multi-dimensional techniques (e.g., NASA-TLX). Therefore, some techniques can only be applied retrospectively. In these cases, it is important that data acquisition is performed as soon as possible after task performance.

Workload ratings may be dissociated from other measures of mental workload (Yeh & Wickens, 1988), so that subjective measures should not be used as the sole basis for assessment.

4.3.2.6 Apparatus Required

Paper-and-pencil rating techniques require minimal technical equipment for data acquisition. However, data acquisition and analysis are sometimes less expensive (and more reliable) if computer-based versions are used, for example with multidimensional scales or those that are based on paired comparisons (SWORD). Computer-based versions are often available or can be generated quite simply.

4.3.3 References

Armstrong Aerospace Medical Research Laboratory (1987). *Subjective Workload Assessment Technique (SWAT): A Users Guide*. Wright-Patterson AFB, OH: Armstrong Aero Medical Research Laboratory.

Endsley, M. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, *37*, 32-64.

Hart, S.G., & Wickens, C.D. (1990). Workload Assessment and Prediction. In: H.R. Booher (Ed.), *MANPRINT. An Approach to Systems Integration* (pp. 257-296). New York: Van Nostrand Reinhold.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: P.A. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-184). Amsterdam: North Holland Press.

Johannsen, G., Moray, N., Pew, R., Rasmussen, J., Sanders, A., & Wickens, C. (1979). Final Report of Experimental Psychology Group. In: N. Moray (Ed.), *Mental Workload, Its Theory and Measurement* (pp. 101-114). New York: Plenum Press.

Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Linton, P.M., Wierwille, W.W., Zaklad, A.L., Bittner, A.C., & Wherry, R.J. (1989). *Operator Workload: Comprehensive Review and Evaluation of Operator Workload Methodologies*. Technical Report 851. Alexandria, VA: U.S. Army Research Institute for Behavioral and Social Sciences.

Matthews, G., Jones, D.M., & Chamberlain, A.G. (1990). Refining the Measurement of Mood: The UWIST Mood Adjective Checklist. *British Journal of Psychology*, *81*, 17-42.

McGuinness, B. (1999). Situational Awareness and the Crew Awareness Rating Scale (CARS). *Proceedings of the 1999 Avionics Conference.* (Report No. ERA Technology Report 99-0815 (Paper 4.3)). 17-18 November 1999. Heathrow, London.

McNair, D.M., Lorr, M., & Droppleman, L.F. (1971). *Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.

NASA (1986). *NASA Task Load Index (TLX): Computerized Version*. Moffet Field, CA.: NASA-Ames Research Center, Aerospace Human Factors Division.

NASA (1986). *NASA Task Load Index (TLX): Paper and Pencil Version*. Moffet Field, CA.: NASA-Ames Research Center, Aerospace Human Factors Division.

O'Donnell, R.D., & Eggemeier, F.T. (1986). Workload Assessment Methodology. In: K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of Perception and Human Performance (42.1-42.49).* Vol. II – Cognitive Processes and Performance. New York: Wiley & Sons.

Pfendler, C., Pitrella, F.D., & Wiegand, D. (1995). *Messung der beanspruchung bei der systembewertung*. FAT-Bericht Nr. 115. Wachtberg: Forschungsinstitut für Anthropotechnik.

Pitrella, F.D., & Käppler, W.-D. (1988). *Identification and Evaluation of Scale Design Principles in the Development of the Sequential Judgement, Extended Range Scale*. FAT-Bericht Nr. 80. Wachtberg: Forschungsinstitut für Anthropotechnik.

Reid, G.B., Bressler, J.R., & Potter, S.S. (1989). *SWAT: A User's Guide*. (Report No. TR-89-023). Wright-Patterson AFB, OH: Armstrong Aerospace Medical Research Laboratory.

Reid, G.B., Shingledecker, C.A., Nygren, T.E., & Eggemeier, F.T. (1981). Development of Multidimensional Subjective Measures of Workload. *Proceedings of the IEEE Int. Conf. on Cybernetics and Society*. 403-406.

Taylor, R.M. (1990). Situational Awareness Rating Technique (SART). The Development of a Tool for Aircrew Systems Design. In: *AGARD Advisory Group for Aerospace Research and Development, Conference Proceedings No. 478*. Neuilly-sur-Seine, France: AGARD.

Vidulich, A.M. (1989). The Use of Judgement Matrices in Subjective Workload Assessment: The Subjective Workload Dominance (SWORD) Technique. *Proceedings of the Human Factors Society 33rd Annual Meeting*. Santa Monica, CA: The Human Factors Society.

Watson, D., Clark, L., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *54*, 1063-1070.

Wierwille, W.W., & Casali, J.G. (1983). A Validated Rating Scale for Global Mental Workload Measurement Applications. *Proceedings of the Human Factors Society 27th Annual Meeting*. Santa Monica, CA: The Human Factors Society.

Wierwille, W.W., & Eggemeier, F.T. (1993). Recommendations for Mental Workload Measurement in a Test and Evaluation Environment. *Human Factors*, *35*, 263-281.

Yeh, Y., & Wickens, C.E. Dissociation of Performance and Subjective Measures of Workload: Measurement of Workload by Secondary Tasks. *Human Factors*, *30*, 111-120.

4.4 THE USE OF MODELS

4.4.1 Fatigue Models

Several models for the prediction of fatigue or alertness exist in the scientific community. Although most models are still under development, they are being applied. All of these models are based on Borbély's two-process sleep model (Daan, Beersma & Borbély, 1984). The differences between the models mainly concern the validation of the models (i.e., the state of validation and the particular target group for the validation). Here, the models are reviewed with respect to their validation and to their particular purpose and application area.

4.4.1.1 Two-Process Model for Sleep Regulation of Borbély

The general structure of the model postulates a single circadian pacemaker, entrained by an external zeitgeber. The pacemaker is presumably located in the SCN (supra-chiasmatic nuclei) of the hypothalamus and normally functions as a single unit. It can be entrained by the light-dark cycle via the retino-hypothalamic tract. Through its efferents it may generate numerous physiological circadian oscillations or synchronize them. Among these are oscillations in the two thresholds, a low threshold (L) and a high threshold (H), for the so-called sleep process (S). S increases monotonically during wakefulness until it reaches H, at which point sleep is initiated. S declines monotonously during sleep until it reaches L, at which point sleep is terminated. The system acts like a thermostat that switches at the thresholds. Since S, H, and L are used as dimensionless variables (i.e., as fractions of the minimum-maximum range of S), the sign of the changes in S is equivocal. We adopt the convention that S decreases during sleep and increases during wakefulness. S may be thought of as a chemical sleep factor, whereas H and L may reflect the sensitivity of hypothetical brain receptors for S. Although these assumptions are not necessary for the present purpose, they facilitate the conceptualization and may eventually lead to a specific neuro-chemical hypothesis.

In the "somnostat" of Borbély's model, the frequency of sleep-wake alternations depends on the interval between the two threshold levels and on the rate of build-up and decrease of S. We assume that external conditions affect the threshold levels. Sleep deprivation experiments are simulated by suspending the upper threshold H, thereby allowing S to increase further. In contrast, bed rest, warmth, darkness, or the absence of social stimulation lowers H so that sleep is facilitated. Culturally determined habits such as naps cause a transitory depression of the upper threshold. The wake threshold L, operative during sleep, seems less influenced by the environment, although the ringing of an alarm clock may be effectively equivalent to a sudden rise in L. The sleep-wake behavior may feed back to the circadian pacemaker. Self-selection of the light-dark cycle should exert an effect on the entrainment of the circadian pacemaker. There is, however, no solid evidence for such feedback. Finally, the sleep-wake cycle may directly affect many physiological oscillations or may exert a masking influence. This should be taken into account when comparing experimental data and theoretical predictions from the model.

4.4.1.2 Factors Causing Fatigue and Included in a "Fatigue Model"

Besides the circadian variation in fatigue and the process S that describes the build-up of fatigue during waking and the recovery during sleep (Daan et al., 1984), two other causes of fatigue are discussed, the time-on-task effect and sleep inertia.

4.4.1.2.1 Time-on-Task Effect

The task itself may contribute to the build-up of fatigue. The more resources the task uses and the longer the task lasts, the larger is its effect on fatigue (i.e., the time-on-task effect). In some way, determining the effect of a particular task is the most difficult aspect of alertness since any given task may produce a different time-on-task effect. A classification of tasks with regard to this effect does not yet exist.

4.4.1.2.2 Sleep Inertia

Sleep inertia defines a period of transitory fatigue and impaired cognitive and sensorimotor performance that follows awakening from sleep (Ferrara & DeGennaro, 2000). The monitoring of several physiological parameters during the period following sleep indicates that the transition to normal waking values is slow. Sleep inertia ceases about 2 to 3 hours after awakening. Therefore, sleep inertia has relevant operational implications.

Sleep inertia is modulated by several factors. The effects of sleep inertia are more pronounced after awakening from slow-wave sleep than from REM sleep. It is likely that sleep inertia interacts with the circadian clock, but there is no confirmation at this time. Sleep duration influences sleep inertia; after very short sleep periods, only a minor sleep inertia is observed.

The time course of sleep inertia effects is such that fatigue and performance initially improve very rapidly. Therefore, the time course may be approximated by an inverse exponential function.

The operational implications of sleep inertia may be serious (e.g., if someone wakes to perform a complex task immediately). In particular, performance on the task will be impaired if the waking occurs from slow-wave sleep.

4.4.1.3 Existing Models

4.4.1.3.1 System for Aircrew Fatigue Evaluation (SAFE)

The SAFE model predicts the level of alertness/fatigue as the sum of two components, one related to the time of day, or more specifically to the circadian rhythm of the individual, and the other to the time since sleep.

The “time-of-day” component represents the diurnal change in alertness from low levels overnight to a peak in the late afternoon. This variation is associated with the internal circadian rhythm or “body clock” and normally remains entrained to the local time of day.

The phase of the circadian rhythm will vary under the influence of time zone transitions or major changes in the sleep-wake pattern. There may also be transient reductions in the amplitude of the rhythm. These changes are modeled by a forced van der Pol equation, the parameters of which have been estimated from data obtained from aircrew on the London to Sidney route (Gundel & Spencer, 1999; Spencer & Robertson, 1999).

The “time since sleep” component contains two separate elements, following the model of Folkard and collegues (Folkard, Akerstedt, Macdonald, Tucker & Spencer, 1999). The first element is the recovery of alertness immediately on waking: the so-called “sleep inertia” effect. The second component is the exponential reduction in alertness associated with increasing time since sleep and the corresponding exponential increase in alertness generated during sleep. This second component is modeled by the so-called “S Process,” which represents the requirement for sleep as a function of the pattern of sleep and wakefulness. The S Process enables the model to estimate the differential influence on alertness of sleep periods of different lengths.

The output from the model consists of levels of alertness on a scale from 0 to 100, where the limits represent the lowest and highest levels that are theoretically achievable. The scale may be converted to the Samn-Perelli alertness scale, which has been validated against (military) air transport operations (Samn & Perelli, 1982).

Most elements of the model have been based on the results of laboratory experiments. However, the ability of the model to predict changes in alertness among aircrew has been established. In this context, the Defence Evaluation and Research Agency, Centre for Human Sciences (DERA CHS) has carried out a comparison between the predictions of the model and subjective levels of alertness on the flight deck provided by the DLR Institute of Aerospace Medicine. These subjective measures had already been found to correlate well with objective measures based on the electrical activity of the brain.

The first comparisons were based on 12 aircrew on the return trip between Düsseldorf and Atlanta and 10 aircrew flying between Hamburg and Los Angeles. The initial results were encouraging. However, agreement was not as good when comparing the results from 22 aircrew flying between Frankfurt and the Seychelles. The outward and return flights were on consecutive nights, with a daytime rest period of about 14 hours, and it is possible that the model may be overestimating the recuperative value of daytime sleep.

The subjective alertness data collected by the DLR in these studies were based on the Samn-Perelli checklist. The analysis of the respective sets of values has enabled a transformation between this scale and the 100-point scale output from the model to be derived. The computer program presents levels of fatigue in a color-coded format based on the Samn-Perelli scale.

4.4.1.3.2 Alert

The model named "Alert" is partly based on the initial DERA (now Qinetiq) model (Spencer & Gundel, 1998). In contrast to that model, it is targeted at fatigue in surface transport. Figure 17 shows an actual shift schedule of a truck driver over 25 days. The shifts last 10:45 hours including a mandatory break of 45 minutes after 4.5 hours of driving. The (grey) sleep periods have been constructed and are not known for this driver. Fatigue including the time-on-task component has been calculated by the model and appears color-coded during driving. Pink and red fatigue values should be avoided, e.g. by introducing breaks and short rest periods.

The model is characterized by the consideration of task-related fatigue. It contains a time-on-task effect that has been validated in a laboratory experiment.

4.4.1.3.3 Sleep, Activity, Fatigue, and Task Effectiveness Model (SAFTE)

The Walter Reed Sleep/Performance Model is based on Borbély's two-process model, as are most fatigue and performance models. It also includes sleep inertia effects on performance. The model differs from other models in that it does not predict fatigue but rather operator performance (Hursh, 1998).

Inputs to the model are the sleep-wake history and the time of day of performance. It is assumed that both factors interact non-linearly in the prediction of performance. The sleep-wake history input contains a function that takes into account that Stage 1 sleep at the beginning of a sleep period or after an arousal does not add to the recuperative value of sleep.

The model parameters were estimated using normalized response speed on the Psychomotor Vigilance Task (PVT) (Dinges, Orne, Whitehouse, & Orne, 1987; Dinges et al., 1997). Validation took place using 66 truck drivers in a laboratory study, Figure 17 (Balkin et al., 2000).

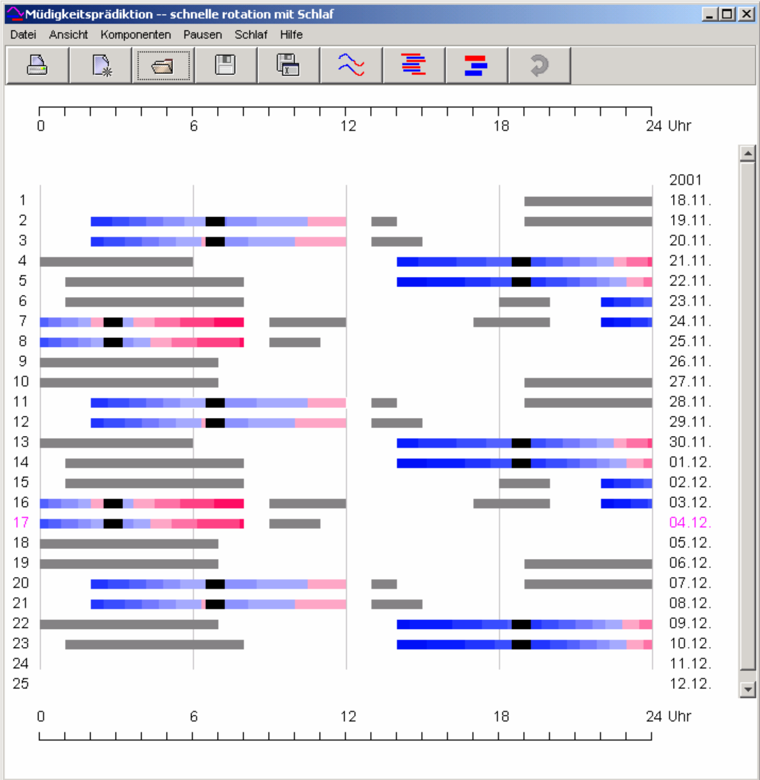

Figure 17: Screenshot of an Actual Shift Schedule of a Truck Driver over 25 Days. The computer program includes a sleep model that was used to construct possible sleep periods (grey). The program predicts fatigue for a given sleep-wake schedule. For duty periods fatigue is color-coded, red indicates fatigue that should be avoided during a driving task.

4.4.1.3.4 Sleep-Wake Predictor

The mathematical/computer model of Akerstedt and Folkard (1997) was developed to predict sleepiness and performance in daily living. The model uses sleep data as input and contains a circadian and a homeostatic component, the latter taking the amount of prior wakefulness and the amount of prior sleep into account. The two components are summed to yield predicted sleepiness as well as performance on monotonous tasks.

The model includes an identification of levels at which performance and alertness impairment start, as well as prediction of sleep onset latency and time of waking from sleep episodes. The intention is to use the model to evaluate work-rest schedules in terms of sleep-related safety risks.

The validity of the model was tested against laboratory and field studies of irregular work hours, using subjective alertness as well as EEG alpha band and theta band power density during waking with eyes open (Akerstedt & Folkard, 1995). Increased alpha and theta activity may be incompatible with adequate perception of visual signals.

The model does not predict task-related fatigue.

4.4.1.3.5 Fatigue Audit Interdyne (FAID)

A model for work-related fatigue has been proposed by Dawson and Fletcher (2001). The only input to this model is the hours of work. Regarding sleep, a statistical distribution of sleep times is assumed and sleep times are not an input to the model.

Fatigue in this model is dependent on the time of day (circadian component). It accumulates during work and dissipates during non-work hours regardless whether there is sleep or not. The effect of work is cumulative and the work-related fatigue lasts for 7 days in the model.

The model has been validated against data, but it is mainly intended for a comparison of work schedules based on the qualitative assumptions made (Fletcher & Dawson, 2001).

4.4.1.3.6 Interactive Neurobehavioral Model

The model output is a linear combination of circadian, homeostat, and sleep inertia components (Jewett & Kronauer, 1999). The effect of light on the circadian component is taken into account but not a possible alerting effect of light.

The model has been validated by laboratory studies in which subjects were exposed to varying light patterns, jet lag, sleep deprivation, and non-24-h schedules (Kronauer, Forger & Jewett, 1999). The model does not have a time-on-task component.

4.4.1.4 Work in Progress

Currently, the relation between fatigue and performance is debated and explored. This is a relatively complicated subject that has many practical implications. Another area that demands further research is the effect of a task on sleepiness and fatigue.

4.4.1.5 References

Akerstedt, T., & Folkard, S. (1995). Validation of the S and C Components of the Three-Process Model of Alertness Regulation. *Sleep*, *18*, 1-6.

Akerstedt, T., & Folkard, S. (1997). The Three-Process Model of Alertness and its Extension to Performance, Sleep Latency, and Sleep Length. *Chronobiology International*, *14*, 115-123.

Balkin, T., Thorne, D., Sing, H., Thomas, M., Redmond, D., Wesensten, N., Williams, J., Hall, S., & Belenky, G. (2000). *Effects of Sleep Schedules on Commercial Driver Performance.* (Report No. DOT-MC-00-133). Washington, DC: U.S. Department of Transportation, Federal Motor Carrier Safety Administration.

Daan, S., Beersma, D.G.M., & Borbély, A.A. (1984). Timing of Human Sleep: Recovery Process Gated by a Circadian Pacemaker. *American Journal of Physiology*, *246*, R161-178.

Dawson, D., & Fletcher, A. (2001). A Quantitative Model of Work-Related Fatigue: Background and Definition. *Ergonomics*, *44*, 144-163.

Dinges, D.F., Orne, M.T., Whitehouse, W.G., & Orne, E.C. (1987). Temporal Placement of a Nap for Alertness. *Sleep*, *10*, 313-329.

Dinges, D.F., Pack, F., Williams, K., Gillen, K.A., Powell, J.W., Ott, G.E., Aptowicz, C., & Pack, A.L. (1997). Cumulative Sleepiness, Mood Disturbance, and Psychomotor Vigilance Performance Decrements during a Week of Sleep Restricted to 4-5 Hours per Night. *Sleep*, *20*, 267-277.

Ferrara, M., & De Gennaro, L. (2000). The Sleep Inertia Phenomenon during the Sleep-Wake Transition: Theoretical and Operational Issues. *Aviation, Space, and Environmental Medicine*, *71*, 843-848.

Fletcher, A., & Dawson, D. (2001). Field-Based Validations of Work-Related Fatigue Model Based on Hours of Work. *Transportation Research Part F*, *4*, 75-88.

Folkard, S., Akerstedt, T., Macdonald, I., Tucker, P., & Spencer, M.B. (1999). Beyond the Three-Process Model of Alertness: Estimating Phase, Time on Shift, and Successive Night Effects. *Journal of Biological Rhythms*, *14*, 577-587.

Gundel, A., & Spencer, M.B. (1999). A Circadian Oscillator Model Based on Empirical Data. *Journal of Biological Rhythms*, *14*, 516-523.

Hursh, S.R. (1998). *Modeling Sleep and Performance within the Integrated Unit Simulation System (IUSS)*. (Technical Report: Natick/TR-98/026L). Final Report for the United States Army Soldier Systems Command; Natick Research, Development and Engineering Center, Natick, MA 01760-5020; Science and Technology Directorate.

Jewett, M.E., & Kronauer, R.E. (1999). Interactive Mathematical Models of Subjective Alertness and Cognitive Throughput in Humans. *Journal of Biological Rhythms*, *14*, 588-597.

Kronauer, R.E., Forger, D.B., & Jewett, M.E. (1999). Quantifying Human Circadian Pacemaker Response to Brief, Extended and Repeated Light Stimuli over the Phototopic Range. *Journal of Biological Rhythms*, *14*, 500-515.

Samn, S.W., & Perelli, L.P. (1982). *Estimating Aircrew Fatigue: A Technique with Implications to Airlift Operations.* (Technical Report No. SAM-TR-82-21). Brooks AFB, TX: USAF School of Aerospace Medicine.

Spencer, M.B., & Gundel, A. (1998). *A PC-Based Program for the Assessment of Duty Schedules in Civil Aviation: The Way Forward.* (Report DERA/CHS/PP5/CR/980069/1.0). Farnborough, UK: DERA.

Spencer, M.B., & Robertson, K.A. (1999). *The Alertness of Aircrew on the London-Sydney Route: Comparison with Predictions of a Mathematical Model*. (Report DERA/CHS/PPD/CR/990261/1.0). Farnborough, UK: DERA.

4.4.2 Operator Performance Modeling

4.4.2.1 Introduction

The complexity of modern military systems is increasing. This greatly increases the amount of information the human operator must process before making decisions and taking action. To achieve knowledge of an operator's actual cognitive needs, models of operator performance must be developed along with reliable and valid methods to assess the central concepts of workload, situational awareness, and operative performance (Angelborg-Thanderz, 1982, 1989, 1997; Svensson, Angelborg-Thanderz, & Wilson, 1999; Svensson, 2000; Svensson & Wilson, 2002).

Operator WorkLoad (OWL) and Operator Performance (OP) have been central concepts for more than thirty years, and the concept of Situational Awareness (SA) has been an actor on the scene for ten to fifteen years. The concepts and their relationships form the basis of research in modelling of operator performance. Models of the operator are needed to describe, and sometimes explain, how the operator copes with situations and the system. The ultimate goal of a model is to reliably predict the outcomes of complex, multi-factorial processes by means of a small number of central concepts.

It has been difficult to formulate operational definitions of the concepts of OWL, SA, and OP, and even harder to develop practical measures. Operational definitions and practical, valid, and reliable measures are, of course, necessary, but this is not enough. Development of models involving the interactions among the concepts, their causal relationships, and how systems, operational factors, and operator experience shape the concepts are a necessary second step. By means of these models, we can predict and estimate the relative sensitivity of OWL, SA, and OP as a function of the complexity of operations, and we can find cognitive and technical “bottlenecks” in the systems. One reason behind the difficulties in the development of useful decision support systems is the lack of useful psychological models of operator performance.

4.4.2.2 Modeling Approaches

4.4.2.2.1 Conceptual Modeling

Modeling can be approached from different angles. Conceptual models are descriptive and they provide a framework for investigation of the components of human performance. They provide a useful technique for examining potential limitations in operator performance. Wickens’ (1992) model of human information processing describes the critical stages of information processing involved in human performance. The model assumes that each stage of processing performs some transformation of the data and requires some time for its operation. Wickens’ (1992) multiple-resource theory is another fruitful example modeling the proposed structure of cognitive processing resources.

4.4.2.2.2 Computer Based Modeling

Another approach is concerned with the development of computer programs that model human performance as well as technical systems. The modeling of technical systems is a prerequisite for and very close to *simulation*. The fidelity of a flight simulator is a function of the validity and reliability of the models of the situation. Detailed and exact data (e.g., physical relationships and algorithms) are generally available for models of technical systems, and, accordingly, the fidelity of such simulations usually is rather high. Recent examples of this kind are threat modeling and the simulation of flight incidents and accidents (Smaili, 2000; Maraoka, & Noriaki, 2000).

Because of the successful modeling of technical systems (e.g., flight and weapons systems), it is tempting to try to model humans in the same way. The modeling of physiological and perceptual processes has already been successful, and these models are now used in the development of simulation systems. However, there is so far an obvious difference between technical/physiological *and* psychological systems with regard to basic knowledge. Even if existing computer models of human cognitive performance seem to have fidelity and validity at first glance, closer inspection often discloses restrictions with respect to their ability to *predict* human behavior and performance. Due to the lack of psychological knowledge, the empirical bases of the models are mostly weak or non-existent.

Despite these shortcomings, cognitive computational modeling can help in characterizing the changes that occur in order to facilitate improved crew performance, because it enables learning and knowledge to be independently and directly manipulated. Models can predict what initial knowledge is required to produce the observed behavior, how new strategies are acquired, and how task knowledge is learned.

Soar (Laird, Newell, & Rosenbloom, 1987) and ACT-R (Anderson, 1993) are the two main symbolic cognitive architectures that can be used to model human behavior. Both approaches reduce much of human behavior to problem solving. Soar does this rather explicitly, being based upon Newell’s information processing theory of problem solving, whereas ACT-R merely implies it by being goal directed.

Predictions of visuomotor tracking behavior with respect to delays can be made using models of human manual control performance. The Crossover Model is frequently used to describe pilot performance (Wickens, 1986). In the crossover model, the operator is represented as a number of simple elements: a gain, a threshold, an information-processing delay, a source of noise, and a filter that can be configured according to the characteristics of the given tracking task.

4.4.2.3 MIDAS (Man-Machine Integrated Design and Analysis System)

MIDAS (Corker, and Smith, 1993; Staveland, 1991, 1994) is a set of software modules and editors that allow simulation of humans interacting with crew station equipment, vehicle dynamics, and a dynamically generated environment. Quantitative models of the operator, the crew stations, and the environment of the vehicle are implemented with emphasis on operator performance under mission conditions. The models of human perception, cognitive behavior, and all responses are detailed and allow analysis of critical areas of human performance such as information management, cognition, and mental workload. MIDAS also allows the inclusion of probabilistic events and errors and is able to model interruption and resumption of tasks in single-operator and multiple-operator settings.

4.4.2.3.1 IPME (Integrated Performance Modeling Environment)

IPME (Dahn, Laughery, and Belyavin, 1997) is an integrated environment of models intended to help analyze human system performance. The base technologies that have been incorporated in IPME are Micro Saint and Human Operator Simulator (HOS). The latter contributes human characteristics to Micro Saint. IPME provides a more or less realistic representation of humans in complex environments, along with interoperability with other model components and external simulations.

4.4.2.3.2 Data-Based Modeling

This modeling approach is based primarily on empirical data, and, accordingly, the resulting models represent the empirical relationships among concepts. The approach is based on “second generation” multivariate statistical techniques that enable statistical tests of causal flow models. These techniques are described in the section on Statistical Techniques included in this volume. Thus, the theory is based on (and can be rejected on the basis of) empirical observations and experience (Jöreskog, & Sörbom, 1984, 1993; Saris; Stronkhorst, 1984).

Causal explanations represent the most fundamental understanding of the processes studied, and such knowledge is invariant over time. It is more important to know that one phenomenon is a cause of another than merely to know that these phenomena appear together. Potentially, knowledge of cause and effect makes it possible to influence reality in an intelligent way.

The techniques are especially suited for non-experimental research and data. The major characteristic of non-experimental research is that the experimenter cannot strictly manipulate the relevant variables. This is often the case in applied research in operational settings (e.g., studies of pilot performance in realistic flight scenarios) (Angelborg-Thanderz, 1989, 1997; Svensson, 1997; Svensson, Angelborg-Thandez, & Sjoeberg, 1993; Svensson, Angelborg-Thanderz, Sjoeberg, & Olsson, 1997; Svensson & Wilson., 2002). The major strength of the technique is that it makes it possible to draw experimental conclusions from non-experimental real and operational situations.

4.4.2.4 References

Anderson, J.R. (1993). *Rules of the Mind.* Mahwah, NJ: Lawrence Erlbaum Associates.

Angelborg-Thanderz, M. (1982). *Assessing Pilot Performance and Mental Workload in Training Simulators*. London: The Royal Aeronautical Society.

Angelborg-Thanderz, M. (1989). *Assessing Pilot Performance in Training Simulators. A Structural Analysis*. In 1989 Spring Convention – Flight Simulation: Assessing the Benefits and Economics. London: The Royal Aeronautical Society.

Angelborg-Thanderz, M. (1997). Military Pilot Performance – Dynamic Decision Making in its Extreme. In F. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision Making Under Stress, Emerging Themes and Applications* (pp. 225-232). Aldershot, Hants, England: Ashgate Publishing Company.

Corker, K.M, & Smith, B.R. (1993). An Architecture and Model for Cognitive Engineering Simulation Analysis: Application to Advanced Aviation Automation (pp. 1079-1088). American Institute of Aeronautics and Astronautics.

Dahn, D.A., Laughery, K.R., & Belyavin, A.J. (1997). *The Integrated Performance Modelling Environment: A Tool for Simulating Human-System Performance*. Proceedings of the 42nd Annual Human Factors and Ergonomics Society Conference (pp. 1037-1041).

Jöreskog, K.G., & Sörbom, D. (1984). *LISREL VI: Analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables, and Least Squares Methods*. Uppsals, Sweden: Department of Statistics, University of Uppsala.

Jöreskog, K.G., & Sörbom, D. (1993). *LISREL8: Structural Equation Modeling with the SIMPLIS Command Language.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Laird, J., Newell, A., & Rosenbloom, P. (1987). SOAR: An Architecture for General Intelligence. *Artificial Intelligence*, *33*, 1-64.

Maraoka, K., & Noriaki, N. (2000). *FBSS-RAIS: Flight Crew Behavior Simulation System – Reconstruction of Accident/Incident Scenario*. AIAA Modeling and Simulation Technologies & Exhibit, Denver, Colorado, August, 2000. American Institute of Aeronautics and Astronautics.

Saris, W., & Stronkhorst, H.P. (1984). *Causal Modelling in Non-Experimental Research. An Introduction to the LISREL Approach.* Amsterdam: Sociometric Research Foundation.

Smaili, M.H. (2000). *Flight Data Reconstruction and Simulation of the 1992 Amsterdam Bijlmermeer Airplane Accident*. AIAA Modeling and Simulation Technologies & Exhibit, Denver, Colorado, August, 2000. American Institute of Aeronautics and Astronautics.

Staveland, L. (1991). *MIDAS TLM: Man-Machine Integrated Design and Analysis System Task Loading Model*. Conference Proceedings IEEE/SMC, 1219-1223.

Staveland, L. (1994). *Man-Machine Integration Design and Analysis System (MIDAS) Task Loading Model (TLM) Experimental and Software Detailed Design Report*. NASA CR-177640.

Svensson, E., (1997). Pilot Mental Workload and Situational Awareness – Psychological Models of the Pilot, In F. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision Making Under Stress, Emerging Themes and Applications* (pp. 261-267). Aldershot, Hants, England: Ashgate Publishing Company.

Svensson, E. (2000). *Models of Pilot Performance – Effects of Motivation and Form.* Paper presented at the 5th Conference on Naturalistic Decision Making, Tammsvik, Stockholm, May 26-28, 2000.

Svensson, E., & Wilson, G.F. (2002). Psychological and Psychophysiological Models of Pilot Performance for Systems Development and Mission Evaluation. *The International Journal of Aviation Psychology*, *12*, 95-110.

Svensson, E., Angelborg-Thanderz, M., & Wilson, G.F. (1999). *Models of Pilot Performance for Systems and Mission Evaluation – Psychological and Psychophysiological Aspects*. AFRL-HE-WP-TR-1999-0215. Wright-Patterson Air Force Base, Ohio: Air Force Research Laboratory.

Svensson, E., Angelborg-Thanderz, M., & Sjöberg, L. (1993). Mission Challenge, Mental Workload and Performance in Military Aviation. *Aviation, Space, and Environmental Medicine*, *64*, 985-991.

Svensson, E., Angelborg-Thanderz, M., Sjöberg, L., & Olsson, S. (1997). Information Complexity – Mental Workload and Performance in Combat Aircraft. *Ergonomics*, *40*, 362-380.

Wickens, C.D. (1986). The Effects of Control Dynamics on Performance. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of Perception and Human Performance* (pp. 39.1-39.60). New York: Wiley.

Wickens, C.D. (1992). *Engineering Psychology and Human Performance*. New York: Harper Collins Publishers Inc.

4.5 STATISTICAL AND MATHEMATICAL TOOLS

4.5.1 Statistical Techniques for Data Reduction and Modeling

4.5.1.1 Background

Correlational statistics, factor analytical techniques, and "second generation" multivariate analytical techniques have been found to be valuable methodological tools in behavioral sciences research (Gorsuch, 1974; Tinsley, and Tinsley, 1987; Hair, Anderson, Tatham, & Black, 1998). Multivariate statistical techniques are important tools for analyzing multiple relationships and application of experimental designs in applied situations. They make possible parsimonious descriptions of complex psychological and physiological relationships and they are prerequisites for modelling of human behavior. By means of 'second generation' multivariate statistics we can analyze causal relationships and the relative effects of different causal factors (Fassinger, 1987; Jöreskog, and Sörbom, 1993).

The techniques are based on correlational statistics i.e. the linear relationships between variables, and the common variance between the variables forms the basis for the analyses. Accordingly, the techniques present the degree of relationship between variables in terms of explained variance. This is more powerful than 'first generation' statistical techniques, which compared group means using t-tests and analyses of variance. Factor analysis (FA) is by far the most widely used data reduction technique, and it forms the basis for related techniques such as cluster analysis, multidimensional scaling, and structural equation modelling.

4.5.1.2 Factor Analysis

4.5.1.2.1 Rationale

Factor Analysis is an analytical technique that reduces a large number of interrelated manifest variables to a smaller number of latent variables or factors. The goal of the technique is to achieve a parsimonious description by using the smallest number of explanatory concepts needed to explain the maximum amount of common variance in a correlation matrix.

By means of psychological and psychophysiological constructs we can reduce and interpret the multitude of human behaviors, and from the empirical relations between the constructs performance models can be developed.

4.5.1.2.2 The Factor Analytical Procedure

The co-variances between variables are the points of departure for FA. The total variance of a variable consists of common, specific, and error variance. Common variance is the co-variance between two or more variables, and specific variance is the reliably measured unique variance of a variable. The objective of FA is to extract the factors behind the common variance. The factor extraction procedures can be divided into *exploratory and confirmative* (hypothesis testing) methods. Explorative solutions cannot be generalized to populations. Generalization requires replications in new samples. LISREL (analysis of linear structural relationships) (Jöreskog, and Sörbom, 1993) is a practical tool for confirmation and generalization of factor structures. LISREL can be used to perform both exploratory and confirmatory FA. LISREL is characterized by two basic components: a structural model and a measurement model. The structural model is a 'path' model, relating independent variables to dependent variables. The measurement model is a maximum likelihood FA defining the relations between manifest variables and latent variables or factors. Above all, the combination of the models offers a powerful method for examination of theories and testing of causal models. The basic principle of FA is to explain as much true variance as possible in the covariance matrix with as few factors as possible. When using confirmative or hypothesis testing FA, the number of factors, and the variables that load on each factor, must be stated prior to the analysis. These techniques test the fit of the data to the hypothesized factor structure. An important tool for factor interpretation is factor rotation. The initial un-rotated factor matrix (a table showing the factor loadings of all variables on each factor) assists in obtaining a preliminary indication of the numbers of factors to extract. Factor rotation results in a more even variance distribution, and in a more interpretable and simpler factor structure. Orthogonal techniques are most preferred on both theoretical and empirical grounds.

4.5.1.3 Multidimensional Scaling (MDS)

4.5.1.3.1 Rationale

MultiDimensional Scaling (MDS) is a procedure for fitting a set of objects or variables in a space (or plane) such that the distances between the objects correspond as close as possible to a given set of similarities or dissimilarities between the objects. Similarities can be measured directly or derived indirectly from correlation matrices (Schiffrin, Reynolds, and Young, 1981; Fitzgerald, and Hubert, 1987).

Usually MDS can fit an appropriate model with fewer dimensions than can FA. Furthermore, MDS provides a dimensional model even if a linear relationship between distances and dissimilarities cannot be assumed. As compared to other multivariate techniques MDS is easy to use and the statistical assumptions are easy to fulfil.

4.5.1.3.2 Procedure

The scaling procedure starts by generating a configuration of points, for which the inter-point distances are a linear function of the input data. From this initial configuration the MDS algorithm constructs better solutions by an iterative procedure. The fit is expressed as a stress value ranging from 0.00 to 1.00. The closer the stress value comes to zero the more adequately the spatial configuration represents the relations between the objects or variables. In contrast to FA no statistical distribution assumptions are necessary, even if some metric conditions must be satisfied.

4.5.1.3.3 Illustrations of the Techniques

Example 1: Factor Analysis

Data from a study by Svensson, and Wilson (2002) will be used to illustrate FA, MDS, and structural equation modelling (SEM). In the study, military pilots answered questionnaires on pilot mental workload (PMWL), situational awareness (SA), and pilot performance (PERF) immediately following simulated air-to-air intercepts. Heart rate (HR) and eye fixation rate (FIXRATE) were registered. The correlations between the five variables were estimated and used as input in a FA.

Figure 18 presents the eigenvalues from the FA extraction procedure. As can be seen two eigenvalues are greater than 1.00 (Kaiser's criterion) and for the other three eigenvalues the error variance dominates the common variance (Cattell's scree test). Our conclusion from the criteria is that a 'two factor' solution is the most parsimonious with respect to proportion of explained common variance.

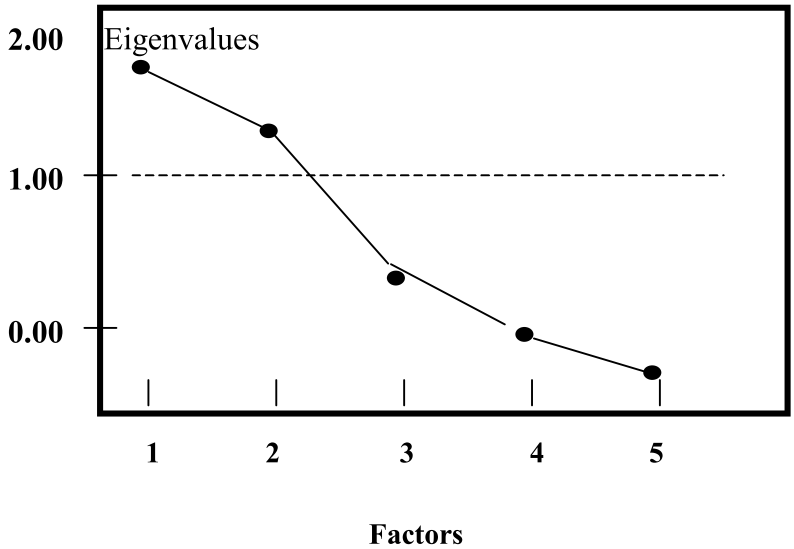

Figure 18: Plot of Eigenvalues Extracted from Successive Residual Correlation Matrices.

Figure 19 presents the factor loadings after varimax rotation. The variables pilot mental workload (PMWL), heart rate (HR), and eye fixation rate (FIXRATE) are significantly loaded in factor 2, and situational awareness (SA) and pilot performance (PERF) are significantly loaded in factor 1. The markers of factor 2 reflect the *mental workload* construct and the markers of factor 1 the *performance* construct. The result illustrates the multifaceted nature of the two constructs. For example, the workload factor is manifest in both the psychological and psychophysiological variables.

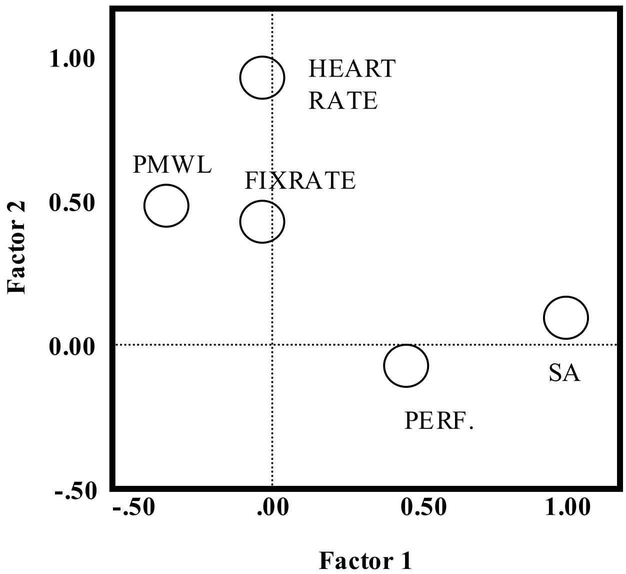

Figure 19: Plot of Loadings for Heart Rate, Pilot Mental Workload (PMWL), Eye Fixation Rate (FIXRATE), Pilot Performance (PERF), and Situational Awareness (SA) for Factors 1 and 2 after Rotation to a Simple Structure.

Example 2: Multidimensional Scaling

The correlation matrix for the five variables was also analyzed by means of MDS. The MDS procedure automatically transforms correlations to dissimilarities. The MDS plot is presented in Figure 20. The fit of the final configuration is perfect and the stress value is .00031. This means that the distances between the variables represent the correlations perfectly in two dimensions (i.e., in a plane).

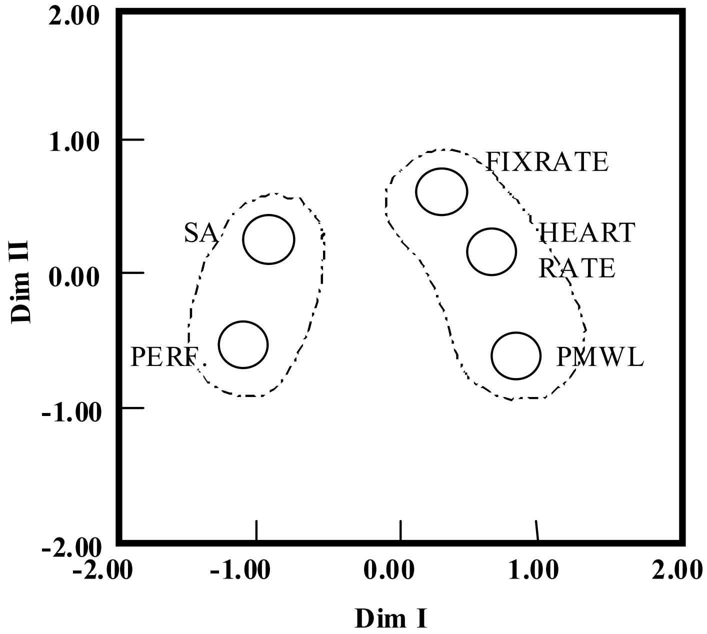

Figure 20: A Two-Dimensional MDS Solution for the Five Variables: Situational Awareness (SA); Pilot Performance (PERF); Eye Fixation Rate (FIXRATE); Heart Rate (HR); and Pilot Mental Workload (PMWL). Stress = .00013.

As can be seen, the dimension I of the MDS solution separates the variables in the same way as the factor solution presented in Figure 20. The second dimension seems hard to interpret but the relative nearness between situational awareness and eye fixation rate appears reasonable.

Example 3: Structural Equation Modelling

The correlations between the five variables of examples 1 and 2 were used as inputs to a structural equation modelling ad modum LISREL. From the FA and MDS analyses we found that the variables formed two factors or dimensions (Figures 19 and 20). The factors were named mental *workload* and *performance*, respectively. Our hypothesized model was that increases in workload cause decreases in the pilots' performance.

The fit of the LISREL solution in Figure 21 is acceptable (Goodness of Fit Index = .85). The ratings of mental workload by means of BFRS, the fixation rate (FIXRATE), and heart rate (HR) are significant markers of the workload factor. This means that an increased activity in the pilot's visual search behavior, an increase in his heart rate, and an increase in his perceived mental workload form a workload factor. The ratings of performance and situational awareness are significant markers of a workload factor. From the solution we can conclude that increases in mental workload cause decreases in the pilots' operative performance (Svensson, Angelborg-Thanderz, & Wilson, 1999).

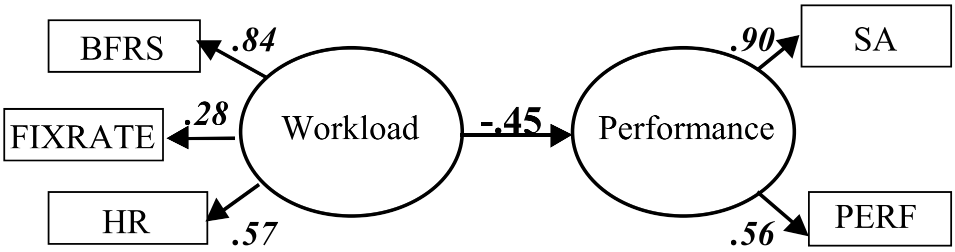

Figure 21: A Structural Model based on the Relationships between Rated Mental Workload using the Bedford Rating Scale (BFRS), Fixation Rate (FR), Heart Rate (HR), Situational Awareness (SA), and Performance Ratings (PERF). Factors are denoted by ellipses and manifest variables by squares. Factor loadings are presented in italics. The effect (-.45) can be considered as a regression or normalized beta weight ranging from -1.00 to 1.00. All coefficients are significant (p < .01).

4.5.1.4 Concluding Remarks

In contemporary research on human behavior when engaged in operating and managing highly complex systems there is a strong demand for data reduction techniques. There is also a need for modelling techniques that are based on empirical data. In this section we have given a brief presentation of the most common techniques with examples showing the development of models of operator functional state and performance.

4.5.1.5 References

Fassinger, R.E. (1987). Use of Structural Equation Modelling in Counselling Psychology Research. *Journal of Counselling Psychology*, *34*, 425-436.

Fitzgerald, L.F., & Hubert, L.J. (1987). Multidimensional Scaling: Some Possibilities for Counselling Research. *Journal of Counseling Psychology*, *34*, 469-480.

Gorsuch, R.L. (1974). *Factor Analysis.* Philadelphia: Saunders Company.

Hair, J.F., Jr., Anderson, R.E., Tatham, R.L., & Black, W.C. (1998). *Multivariate Data Analysis.* New Jersey: Prentice Hall.

Jöreskog, K.G., & Sörbom, D. (1993). *LISREL8: Structural Equation Modeling with the SIMPLIS Command Language.* Hillsdale : Lawrence Erlbaum Associates, Inc.

Schiffrin, S.S., Reynolds, M.L., & Young, F.W. (1981). *Introduction to Multi-Dimensional Scaling Theory, Methods, and Applications*. New York: Academic Press.

Svensson, E., Angelborg-Thanderz, M., & Wilson, G.F. (1999). Models of Pilot Performance for Systems and Mission Evaluation – Psychological and Psychophysiological Aspects, AFRL-HE-WP-TR-1999-0215. Wright-Patterson Air Force Base, OH: Air Force Research Laboratory.

Svensson, E., & Wilson, G.F. (2002). Psychological and Psychophysiological Models of Pilot Performance for Systems Development and Mission Evaluation. *The International Journal of Aviation Psychology*, *12*, 95-110.

Tinsley, H.F., & Tinselly, D.J. (1987). Uses of Factor Analyses in Counselling Psychology Research. *Journal of Counselling Psychology*, *34*, 414-424.

Wilkinson, L. (1990). *SYSTAT: The System for Statistics*. Evanston: SYSTAT Inc.

4.5.2 Artificial Neural Networks

4.5.2.1 Background

Artificial neural networks (ANNs), more commonly referred to as simply neural networks (NN), are made up of interconnected nodes or units. The network of identical connected nodes is usually implemented in software, but dedicated analog/digital hardware units are often used in commercial and military products. NNs are used in a wide variety of signal processing, pattern recognition, and feature extraction applications. Their relative simplicity of design and their ability to perform nonlinear data processing lead to many applications in military, medical, and commercial systems.

Back-propagation is the most commonly implemented ANN algorithm and is used in roughly ninety percent of all applications. A back-propagation neural network classifier maps input vectors to output vectors in two phases. First, the network learns the input-output classification from a set of training vectors. Then, after training, the network acts as a classifier for new vectors.

The back-propagation algorithm initializes the network with a random set of weights for each fully connected layer, then the network trains using the input-output pairs. The learning algorithm uses a two-stage process for each pair: forward pass and backward pass. The forward pass propagates the input vector through the network until it reaches the output layer. First, the input vector propagates to the hidden units. Each hidden unit calculates the weighted sum of the input vector and its associated interconnection weights. Each hidden unit uses the weighted sum to calculate its activation. Next, hidden unit activation propagates to the output layer. Each node in the output layer calculates its weighted sum and activation. Figure 22 shows the forward pass and Figure 23 is a typical unit featuring the summation and the activation. The output of the network is compared to the expected output of the input-output pairs; and their difference defines the output error. In the second stage of network training, the output error propagates backward to update the network weights. First, the error passes from the output layer to the hidden layer updating output weights. Next, each hidden unit calculates an error based on the error from each output unit. The error from the hidden units updates the input weights. One training epoch passes when the network processes all the input-output pairs in the training set. Training stops when the sum-squared error is acceptable or when a predefined number of epochs is executed. The algorithm (backward pass) attempts to minimize the error or energy function defined by:

$$E = \sum_{i=1}^{m} \left\| \bar{z}_i - \bar{t}_i \right\|^2 , \qquad (1)$$

where *m* is the size of the training set, *z* is the neural network output vector, and *t* is the expected output for each training input-output pair *i*.

It may be simpler to examine the algorithm as a series of steps. The steps for implementing a back-propagation neural network are as follows (Lippmann, 1987):

- Initialize the weights (w_i) and biases (b_i), where *i* is the current iteration.
- Present the input matrix (p) and the target vector (t).
- Calculate the output of the network (z_i).
- Calculate the error ($e = z_i - t$).
- Determine the new weights (w_{i+1}) where $i+1$ is the next iteration.
- Determine the new learning rate.
- Repeat steps 2 through 5 until desired error limit is achieved.

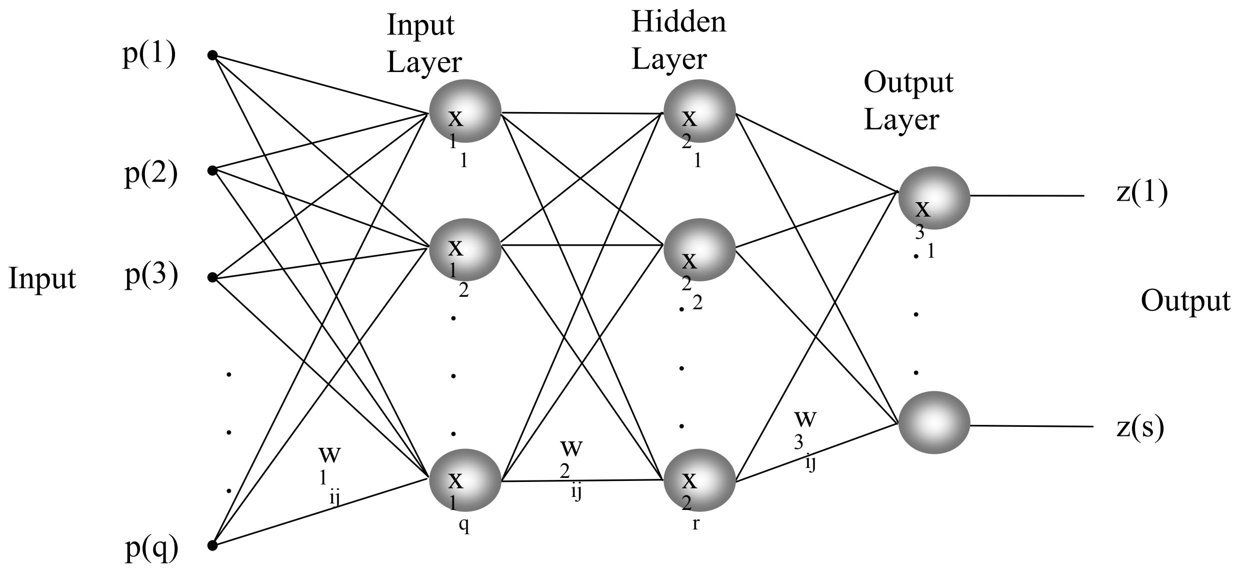

Figure 22: Network Architecture showing a Fully Connected Network with a Number of Neurons in each Layer. The form of the logistic sigmoid activation function is provided.

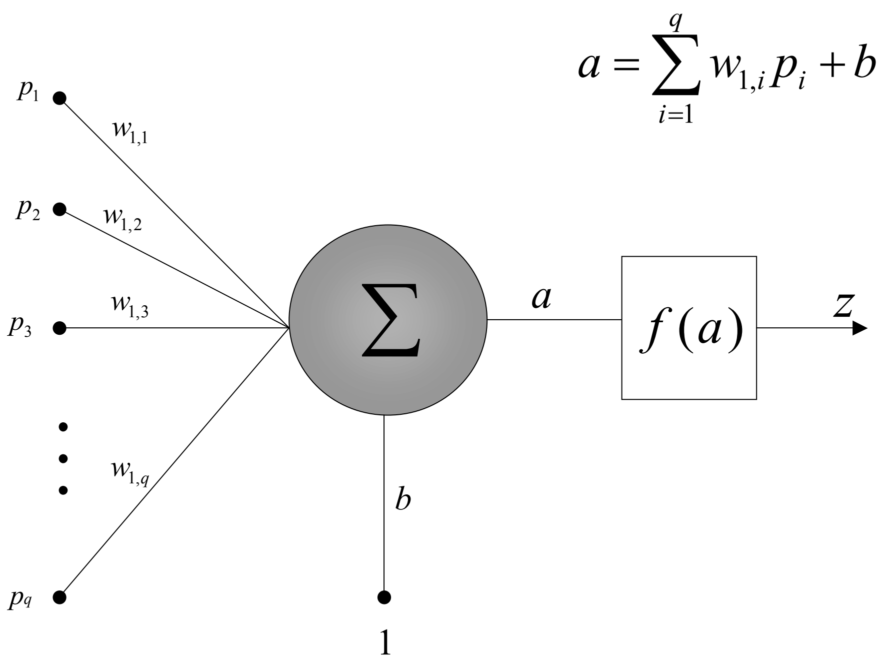

Figure 23: Individual Neuron showing the Weighted Sum of the Inputs followed by the Logistic Sigmoid Activation Function, *f (a)*.

Mathematically, these steps were as given below (Haykin, 1999; Widrow and Stearns, 1985; Widrow and Lehr, 1990). The weights and biases are initialized usually using a random number generator and limiting the values to the range –0.5 to 0.5, which is the nearly linear region of the hyperbolic sigmoid activation function.

The output of the network is determined by propagating the normalized input through each layer of the back-propagation neural network. It is necessary to examine the output of an individual neuron and then expand that understanding to the framework of the entire network. As shown in Figure 23, the output of the individual node or neuron is

$$z = f(a) \tag{2}$$

and

$$a = \sum_{j=1}^{q} \left(w_{1j} p_j + b \right), \tag{3}$$

where w_{1j} is the weight, p_j is the input, b is the bias and $f(a)$ is the activation function acting on a. The figure suggests that this neuron is in the input layer since the leading index on the weight is 1. Generalizing to any neuron results in

$$z_j = f(a_j) \tag{4}$$

and

$$a_j = \sum_{j=1}^{q} \left(w_{ij} p_j + b_j \right). \tag{5}$$

Activation functions can be linear or nonlinear. A common activation function is a sigmoidal nonlinearity. In our case, it is a logistic sigmoid function with an output range $0 \le f(a) \le 1$ in the form

$$f(a) = \frac{1}{1 + e^{-a}}. \tag{6}$$

The error is simply the difference between the output of the network and the expected target value:

$$E_k = \sum_{i=1}^{s} (z_i - t_i)^2, \tag{7}$$

where k is the error for the current input exemplar.

We can adjust the weights and try to minimize the error E_k through the backward path. Although the activation function is nonlinear, it is differentiable and we can compute $\frac{\partial E_k}{\partial w_{ij}}$, which we will use in our selection of a learning rule. The network algorithm is an extension of the Widrow-Hoff learning rule (Widrow and Lehr, 1990), which is a gradient descent algorithm based on Widrow's earlier work in Adaline and Madaline neural networks. This rule adjusts the weights using a steepest descent algorithm,

$$w_{ij}(n) = w_{ij}(n-1) - \mu \frac{\partial E}{\partial w_{ij}}, \tag{8}$$

where μ is a constant that controls the speed of convergence (learning rate).

Adaptive learning and momentum were used to decrease the time required for training the networks and to ensure the network reaches a global minimum. Typically, gradient descent methods use a fixed learning rate to control the rate of convergence. However, it is difficult to determine an optimum rate. If the fixed learning rate is too large, the gradient descent algorithm becomes unstable due to oscillations. If the learning rate is too small, the incremental steps along the error surface are small and in turn the algorithm takes a long time to converge to the desired error. Adapting the learning rate to optimize the learning progress can maintain stability while keeping the learning rate as large as possible to improve the rate of convergence. As the slope of the local error surface increases, the learning rate decreases to control stability.

Momentum prevents the network algorithm from becoming trapped at a local minimum. Essentially, the algorithm will “jump over” or ignore small perturbations in the error surface. Modification of the delta-learning rule to include momentum results in a new learning rule

$$w_{ij}(n) = \alpha w_{ij}(n-1) - \mu \frac{\partial E}{\partial w_{ij}}, \tag{9}$$

where α is the momentum and μ is the learning rate.

This process is repeated until a desired error limit is achieved. The desired error limit is problem-specific and must be determined. Once trained, network weights are fixed and the net acts as a pattern classifier. As a classifier, the network examines input vectors it has never seen and predicts the class of the input vector.

4.5.2.2 State of the Art

There are numerous variations on the design of NNs, depending on the application. Feedforward, multilayer (usually three) is the basic design used in the majority of applications. The NN developer can adjust the number of layers, the number of nodes in each layer, the activation function in each node, the number and pattern of interconnections between the nodes in each layer, and the weight adjustment rules and algorithms that are used during network training. NNs can be trained using supervised learning or the back-propagation method where the input pattern is applied to the NN and the output is calculated. If the desired output is obtained, then learning is complete. If the desired output is not obtained, the interconnection weights are adjusted to minimize the error between the actual and desired outputs, and the process is repeated until the required output is obtained for a given input. A developer can utilize a wide variety of learning rules that define how the weights are adjusted between each pass of the data. In unsupervised learning, nodes in a layer can inhibit other nodes in the layer via additional connections within the layer. The node with the highest activity inhibits the other nodes in the layer from generating any output. Both supervised and unsupervised training of the NN can require an inordinate amount of effort. NNs based on more realistic models of actual nerves and central nervous system neuronal networks are being actively researched.

In the fields of both cognition and physiology, neural networks have been used for data fusion, noise reduction, peak detection, waveform analysis, and data classification.

There are numerous research papers, textbooks, web sites, and magazine articles written on the subject of neural networks, and their applications to many fields, some of which are cited in the reference list. The papers by Krogmann (1997) and Collins (1997) in the AGARD Lecture Series on Advances in Soft-Computing Technologies and Application in Mission Systems provide both an excellent overview of neural networks and a detailed mathematical description of neural networks and control.

4.5.2.2.1 Where Can Neural Nets Be Used

As in the case of fuzzy logic techniques, neural networks can provide a concise and very robust description of operator functional state, and, as in the case of fuzzy logic, a wide variety of variables can be input into the neural network. Unlike fuzzy logic and linear statistical techniques, it is often impossible to understand how or why the NN is producing a given result.

4.5.2.2.2 When Are Neural Nets Appropriate

Neural networks are basically a nonlinear extension to traditional statistical techniques and are often treated as another technique in the arsenal of statistical modeling methodologies. NNs are appropriately used in signal processing (e.g., noise reduction), and pattern and feature recognition (e.g., ECG waveform identification, image analysis). Formal models and mappings using more established statistical techniques may be more appropriate or just as powerful.

4.5.2.3 General Advantages/Disadvantages

Neural networks provide a very powerful technique for signal processing and pattern recognition. They can be over-trained with the learning set of data, such that when presented with a new data set fail to perform as expected. Since it is often impossible to understand how or why the NN is producing a given result, the neural net may be trained to recognize a common feature of the data that is not the feature of interest. The NN developer must provide a robust set of training data.

The number of hidden units required is usually not known. Hidden units are the key to network learning and force the network to develop its own internal representation of the input space. The network that produces the best classification with the fewest units is selected as the best topology. A net with too few hidden units cannot learn the mapping to the required accuracy since the smaller hidden layer would limit interaction of the input space. Too many hidden units allow the net to “memorize” the training data and the net will not generalize well to new data.

4.5.2.4 Software Required

There are numerous software packages and add-on packages to existing statistical and numerical analysis packages available to support neural network design and data analysis, running under both the Windows and Unix environments. Lists of freeware, shareware, and commercial software packages and programs are listed on Web sites dedicated to neural networks. The data manipulation and statistical capabilities of these packages allow for comparison of neural networks to other analysis techniques.

4.5.2.5 Personnel Required

The development of neural networks for a specific application requires an in-depth knowledge of the technology, as well as understanding of alternative statistical techniques. Once a network is constructed, the very nature of the technique allows the use of a network by others who where not involved in the design or the training.

4.5.2.6 References

Collins, D.J. (1997). Introduction to Neural Networks Basic Theory and Application Potential for Mission Systems. In *Advances in Soft-Computing Technologies and Application in Mission Systems*. AGARD Lecture Series 210. Neuilly-sur-Seine, France: AGARD.

Krogmann, U. (1997). Introduction and Overview. In *Advances in Soft-Computing Technologies and Application in Mission Systems*. AGARD Lecture Series 210. Neuilly-sur-Seine, France: AGARD.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*, Upper Saddle River, NJ: Prentice Hall.

Lippmann, R.P. (1987). An Introduction to Computing with Neural Nets. *IEEE Acoustics, Speech, and Signal Processing Magazine*, *4*, 4-22.

Widrow, B., & Lehr, M. (1990). 30 Years of Adaptive Neural Networks: Perceptron, Madaline and Backpropagation. *Proceedings of IEEE*, *78*, 1415-1442.

Widrow, B., & Stearns, S.D. (1985). *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall.

4.5.3 Fuzzy Logic

4.5.3.1 Description of Fuzzy Logic

L.A. Zadeh (1965) introduced fuzzy logic (or fuzzy set theory) specifically to formalize the representation, and more importantly, management of imprecise or approximate knowledge. One important critical feature of fuzzy logic in its application to operator assessment is that allows for a graded or partial membership in a given class, and *membership in more than one class*.

4.5.3.2 State of the Art

In the fields of both cognition and physiology, statistical techniques, including neural networks, have been the method of choice in addressing the problem of data fusion. In engineering and medicine, techniques to combine and summarize complex electro-mechanical systems and patient states, based on the fundamentals of fuzzy logic and fuzzy inference, have become increasingly popular in the last 20 years (Martin, 1994; Cox, 1992; de Silva, 1995; Mendel, 1995). Despite significant distrust and misunderstanding as to what fuzzy logic was or was not, in engineering disciplines the extension of fuzzy logic to fuzzy control has gained wide acceptance in many areas, including safety critical systems and consumer products.

There are numerous research papers, textbooks, web sites, and magazine articles written on the subject of fuzzy logic, and its application to many fields, some of which are cited in the reference list. The papers by Krogmann (1997) and Bouchon-Meunier (1997) in the AGARD Lecture Series on Advances in Soft-Computing Technologies and Application in Mission Systems provides both an excellent overview of fuzzy logic and a detailed mathematical description of fuzzy logic, inference, and control.

4.5.3.2.1 Where Can Fuzzy Logic Be Used

Fuzzy logic techniques can provide a concise and very robust description of operator functional state. Once one is familiar with the language and process of the fuzzy logic approach it is very easy to understand why the analysis is producing the result it does, unlike some statistical and neural network techniques. Additional variables, membership functions, and expert rules can be readily incorporated into a fuzzy expert system.

It is straightforward to manipulate the membership functions of both the fuzzification and defuzzification process, i.e., the analyst can “play” with the membership functions, the fuzzification rules, and the defuzzification process until the functional metric “makes sense”. Fuzzy logic techniques can be combined with other artificial intelligence technologies, such as neural nets and genetic algorithms, and standard statistical techniques can be used to define and adjust the membership functions.

In spite of the extensive literature on the theory and applications of fuzzy logic, there is almost no literature on the use of the technique in the analysis of cognitive and physiological data to assess performance or operator state, except for some very preliminary work on aircrew performance assessment (Fraser, 1998).

With the development and refinement of multiple physiological and performance based measures, the problem of calculating a single metric that defines or describes the overall operator functional state is becoming more problematic. The problem can be described as:

> *given a set of **n** independent or dependent measures of operator state, such as heart rate, reaction time, and eye-blink rate, is there a single value or metric that incorporates the information from each measurement, and accurately reflects the **overall** functional state of the operator.*

Basically this is a problem of sensor fusion – common in many civilian and military applications, especially image data processing.

The membership function for a set maps each element of the set to a membership value between 0 and 1 and uniquely describes that set. The values of 0 and 1 describe "not belonging to" and "belonging to" a conventional set respectively. Values in between represent "fuzziness". Determining the membership function is subjective to varying degrees depending on the situations. It depends on the expert's perception of the data in question, but does not depend on randomness. This distinguishes fuzzy set theory from probability and statistical theory. Such an approach allows us to avoid the problem of hard limits, e.g., all heart rates below 45 bpm are slow, and thus we would define a heart rate of 46 bpm as not slow. In fuzzy logic we would say individuals with a heart rate of 46 bpm would have a certain degree of membership in the "slow" heart rate set, and a certain degree of membership in the "not slow" heart rate set.

A second important feature of fuzzy logic is that information from heterogeneous data types can be easily manipulated and combined. Both physiological data, performance metrics, and observer's perceptions can be formalized in terms of membership functions and degree of membership. The data can be integrated into a set of expert rules. For example:

> *if heart rate variability is low, and EEG delta wave power is high, and subject response time is very slow, and the subject appears to be asleep, then subject's arousal level is low.*

A fuzzy logic approach to both inference, and control is basically a rule based system, using the expert knowledge of the domain specialist as in traditional expert systems, incorporating fuzzy set theory to accommodate the inherent "fuzziness" of the linguistic labels applied by the domain expert, i.e., the heart rate is *slow*. From a functional state assessment point of view, a fuzzy logic approach has the inherent result of producing a single metric from multiple state measures. Since the fuzzy logic approach is an extension of an expert based rule approach, it incorporates the expertise and intuitive knowledge of the analyst. An example of the application of fuzzy logic to operator state assessment is the development of a Pilot State Estimator (PSE) of the physiological state of a tactical pilot exposed to high Gz levels, where ECG and head-level pulse waveforms can be monitored in real-time. The possible PSE or output values of the fuzzy logic analysis are:

- green, indicating good physiological condition
- yellow, indicating a compromised state
- red, indicating severe compromise

The inputs to the algorithm are the acceleration of the aircraft (Gz), head-level blood pressure pulse amplitude, and the delay between the R wave of the ECG waveform and the arrival of pulse wave at head-level. The PSE is computed using fuzzy logic in three stages:

1. The "fuzzification" of crisp values, accomplished by evaluating the values against the applicable membership sets.

2. The application of the fuzzy values to a fuzzy rule set, which will produce a set of membership values.
3. The defuzzification of the set of values to produce a single crisp value, which can be refuzzified to produce a single pilot state estimate of red, yellow, or green.

Gz and pulse amplitude are evaluated against membership functions defining low, medium and high sets. In the case of pulse wave delay, short, medium and long. Figure 24 shows the membership functions for the Gz parameter. If the aircraft is at 7 Gz, the pilot has a 0.3 membership in the high Gz set and a 0.7 membership in the medium Gz set, i.e., he has some degree of exposure to both medium and high Gz – hence the fuzziness. Each of the other parameters will have a unique set of membership functions. Once the membership values have been computed they are applied to a collection of rules that defines the membership function for the PSE in terms of the membership values of the other two parameters. These rules are typically posed in a sentence, for example:

If Gz is high, pulse amplitude is high and pulse wave delay is short, then PSE is green.

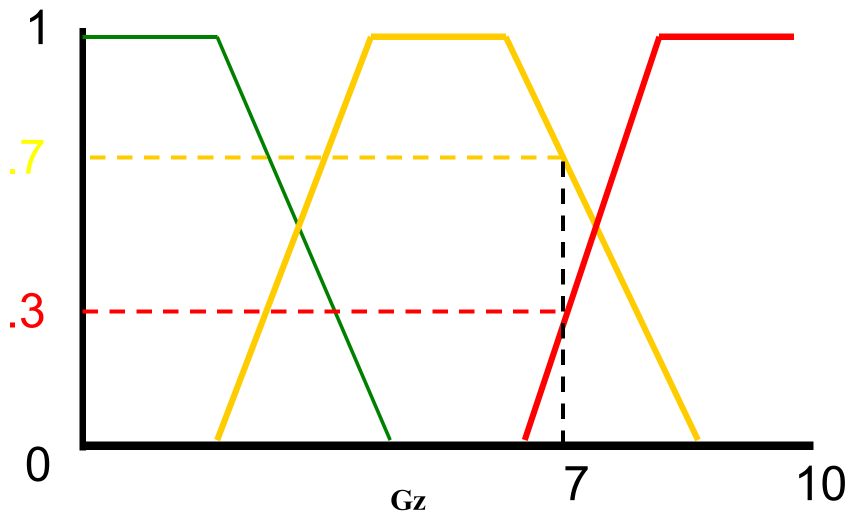

Figure 24: Fuzzy Logic Membership for the Gz Parameter. If the pilot is at a Gz of 7, his membership in the high Gz set (red) is 0.3, and his membership in the medium Gz set (yellow) is 0.7.

The membership value of the PSE for each rule is equal to the *minimum* weighting factor of the three input parameters. This is a logical *and*ing of the values. So if at some point in time the pilot belongs to the high Gz set with a membership of 0.3 in the high (Figure 24), and the pulse amplitude is high (0.5) and the pulse wave delay is short (0.4), then the PSE computed by the above rule is green with a value of 0.3 (the minimum of the three values). So, at the end of this step we have a collection of membership values for PSE, one for each rule that fired.

Once the rules have been applied, we need to convert the set of membership values into a single crisp value, from which we derive our final fuzzy value. We do this using another set of membership functions describing the range of output or PSE values (Figure 25). In the example, we have fired three rules which have given a PSE of green (0.3), yellow (0.2) and red (0.1). We are interested in the area of each trapezoid which is at or below the membership value. We compute the centroid of the three combined trapezoids. The centroid of this area gives the crisp value for the PSE, in this case, 0.65. We can "refuzzify" this value by applying it to the three membership functions in the same manner as in step one, and in this case obtain membership values of red (0.0), yellow (0.65) and green (0.25). Our last rule is that the final value of the

PSE is equal to that set which has the largest weighting, in this case PSE is “yellow” since its membership value was the highest.

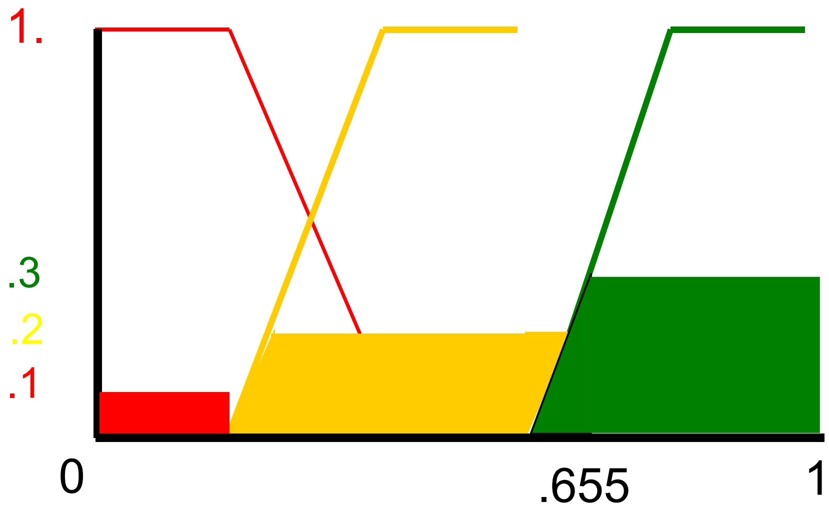

Figure 25: Fuzzy Logic Membership for the PSE Output Parameter. Three rules have fired giving outputs of 0.1, 0.2, and 0.3. These are mapped on to the fuzzy sets of the PSE and the centroid of the combination of the selected areas of each membership function is calculated.

4.5.3.2.2 When is Fuzzy Logic Inappropriate

Fuzzy logic is an alternative approach to statistical analysis and is not based on any probability theory. It is not appropriate for development of formal statistical models of the data such as regression, anova, and general linear modeling. In those cases where only a single metric is collected, fuzzy logic would not be appropriate. Fuzzy logic is not a replacement for statistical tools or neural networks. Fuzzy logic can be combined with more traditional statistical techniques, using statistical analysis to identify membership functions, or in analyzing the defuzzified composite metrics arising from the firing of the fuzzy rule base.

4.5.3.3 General Advantages/Disadvantages

Fuzzy logic is a very powerful technique for the integration of multiple metrics into the estimation of single number to describe operator state. It provides a different, and potentially more useful approach in addressing our particular problem, i.e., what is the “state of the individual operator”. Not “what is the probability that the operator is in a particular state”, or “what is the distribution of states for all the operators”. As in the case of other expert systems, application of fuzzy logic requires a *domain* expert for each of the parameters used in the fuzzy rule base. The domain expert(s) must be conversant with the underlying theory of fuzzy logic and the application development environment. Although a very robust technique and a powerful data reduction technique, extensive testing and sufficient test data is required in order to optimize the membership functions for both input and output variables.

4.5.3.4 Software Required

There are a large number of software packages and add-on packages to statistical packages available to support fuzzy logic data analysis, running under Windows and Unix environments. Lists of freeware, shareware, and commercial software packages and programs are listed on Web sites dedicated to fuzzy logic. The data manipulation and statistical capabilities of these packages allow for comparison of fuzzy logic to other analysis techniques.

4.5.3.5 Personnel Required

Fuzzy logic analysis is based on integrating domain knowledge and a good understanding of the steps involved in fuzzification, expert rule development, and the defuzzification process. A significant amount of information and judgement must be collected from a range of domain experts in order to develop the membership functions for each of the input variables and the output state variable, as well as to itemize the expert system rules. No knowledge of statistical techniques or statistical software is required.

4.5.3.6 References

Bouchon-Meunier, B. (1997). Fundamentals of Fuzzy Logic, Fuzzy Inference, and Fuzzy Control. In *Advances in Soft-Computing Technologies and Application in Mission Systems*: AGARD Lecture Series 210. Neuilly-sur-Seine, France: AGARD.

Cox, E. (1992). Fuzzy Fundamentals. *IEEE Spectrum*, October: 58-61.

de Silva, C.W. (1995). *Intelligent Control and Fuzzy Logic Applications*. Boca Raton, Florida: CRC Press.

Fraser, W.D. (1998). *Human Factors of CC-130 Operations: Future Aircraft Performance Measures.* Technical Report, R-98-16. Toronto: DCIEM.

Krogmann, U. (1997). Introduction and Overview. In *Advances in Soft-Computing Technologies and Application in Mission Systems*: AGARD Lecture Series 210. Neuilly-sur-Seine, France: AGARD.

Martin, J.F. (1994). Fuzzy Control in Anesthesia. *Clinical Monitoring*, *10*, 77-80.

Mendel, J.M. (1995). Fuzzy Logic Systems for Engineering: A Tutorial. *IEEE 83*, 345-377.

Zadeh, L.A. (1965). Fuzzy Sets. *Information and Control*, *8*, 338-353.

4.5.4 Independent Components Analysis (ICA) of Complex Psychophysiological Data

4.5.4.1 Description of ICA

Over the past several years, blind source separation by Independent Component Analysis (ICA) has received significant interest because of its potential application in signal processing such as in speech recognition systems, telecommunications and medical signal processing. ICA is designed to extract independent signal sources given only sensor observations that are unknown linear mixtures of independent source signals. Essentially, ICA is a way of determining a linear non-orthogonal coordinate system in multivariate data. The directions of the axes of this coordinate system are determined by both the second and higher order statistics of the original data. The goal is to perform a linear transform which makes the resulting variables as statistically independent from each other as possible.

4.5.4.2 Background

The development of ICA has its roots in the classic signal processing problem of separating mixed signal sources observed in an array of sensors. Seminal work on blind source separation was performed by Herault and Jutten (1986), who introduced an adaptive algorithm in a simple feedback architecture that was able to separate several unknown independent sources. Unfortunately, the results of their algorithm were poorly understood and led to Comon's paper (1994) defining the problem, and to his solution using fourth-order statistics. Much work took place in this period in the French signal processing community, including Pham, Garat, and Jutten's (1992) Maximum Likelihood approach that subsequently formed the basis of Cardoso and Laheld's (1996) EASI method. Bell and Sejnowski (1995) put the blind source separation problem into an elegant information-theoretic framework and demonstrated the separation and deconvolution of mixed sources, providing a foundation for the application of ICA to psychophysiological and other data types described in the subsequent paragraphs. Below, we sketch the derivation and development of Infomax ICA.

Mathematically, the ICA problem is as follows: We are given a collection of **N**-dimensional random vectors, **x** (sound pressure levels at **N** microphones, **N**-pixel patches of a larger image, outputs of **N** scalp electrodes recording brain potentials, or nearly any other kind of multi-dimensional signal). Typically there are diffuse and complex patterns of correlation between the elements of the vectors. ICA, like Principal Component Analysis (PCA), is a method to remove those correlations by multiplying the data by a matrix as follows:

$$\mathbf{u} = \mathbf{W}\mathbf{x} \tag{1}$$

(Here, we imagine the data is zero-mean --- see below for preprocessing details.). But while PCA only uses second-order statistics (the data covariance matrix), ICA uses statistics of all orders and pursues a more ambitious objective. While PCA simply decorrelates the outputs (using an orthogonal matrix **W**), ICA attempts to make the outputs statistically independent, while placing no constraints on the matrix **W.** Statistical independence means the joint probability density function (p.d.f.) of the output *factorizes*:

$$p(\mathbf{u}) = \prod_{i=1}^{N} p_i(u_i) \tag{2}$$

while decorrelation means only that $\langle \mathbf{u}\mathbf{u}^T \rangle$, the covariance matrix of **u**, is diagonal (here <> means average).

Another way to think of the transform in (1) is as

$$\mathbf{x} = \mathbf{W}^{-1}\mathbf{u} \tag{3}$$

Here, **x** is considered the linear superposition of *basis functions* (columns of $\mathbf{W}^{-1}$), each of which is activated by an independent component, $\mathbf{u}_i$. We call the rows of **W** *filters* because they extract the independent components. In orthogonal transforms such as PCA, the Fourier transform and many wavelet transforms, the basis functions and filters are the same (because $\mathbf{W}^T = \mathbf{W}^{-1}$), but in ICA they are different.

A more general linear transform of **u** is the affine transform: $\mathbf{u} = \mathbf{W}\mathbf{x}+\mathbf{w}$ where **w** is an N-by-1 'bias' vector that centers the data on the origin. If we assume the independent component p.d.f.'s, $p_i(u_i)$ are roughly symmetrical, then it is simpler to subtract the mean, $\langle\mathbf{x}\rangle$, from the data beforehand. A second preprocessing step that speeds convergence is to first 'sphere' the data by diagonalizing its covariance matrix:

$$\mathbf{x} \leftarrow 2\left\langle \mathbf{x}\mathbf{x}^T \right\rangle^{-1/2}\left(\mathbf{x} - \langle \mathbf{x} \rangle\right) \tag{4}$$

This yields a decorrelated data ensemble whose covariance matrix satisfies $\langle\mathbf{x}\mathbf{x}^T\rangle=4\mathbf{I}$ where **I** is the identity matrix. This is a useful starting point for ICA decomposition. This sphering method is not PCA but rather zero-phase whitening which constrains the matrix **W** to be symmetric. By contrast, PCA constrains it to be orthogonal, and ICA, also a decorrelation technique but without constraints on **W**, finds its constraints in the higher-order statistics of the data.

The objective of the Infomax ICA algorithm is to minimize *redundancy* between the outputs. This is a generalization of the mutual information:

$$I(\mathbf{u}) = \int p(\mathbf{u}) \log \frac{p(\mathbf{u})}{\prod_{i=1}^{N} p_i(u_i)} d\mathbf{u} \tag{5}$$

This redundancy measure has value 0 when the p.d.f. **p(u)** factorizes, as in (2), and is a difficult function to minimize directly. The insight that led to the Infomax ICA algorithm was that **I(u)** is related to the joint entropy, **H(g(u))**, of the outputs passed through a set of sigmoidal non-linear functions, g_i :

$$I(\mathbf{u}) = -H(\mathbf{g}(\mathbf{u})) + E\left[\sum_i \log \frac{|g'_i(u_i)|}{p_i(u_i)}\right] \tag{6}$$

Thus, if the absolute values of the slopes of the sigmoid functions, $|g'_i(u_i)|$ are the same as the independent component p.d.f.'s, $p_i(u_i)$ then Infomax (maximizing the joint entropy of the **g(u)** vector), will be the same as ICA (minimizing the redundancy in the **u** vector).

The principle of 'matching' the g'_i's to the p_i's is illustrated in Figure 26, where a single Infomax unit attempts to match an input Gaussian distribution to a logistic sigmoid unit, for which:

$$g(u) = \frac{1}{1+e^{-u}} \tag{7}$$

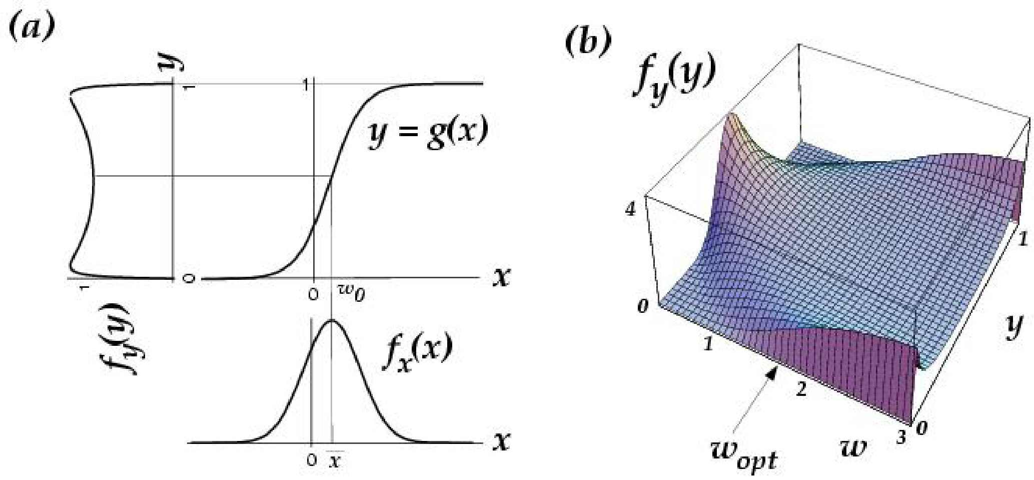

Figure 26: Optimal Information Flow in Sigmoidal Neurons (*left*). Input **x** having probability density function *p(x)*, n this case a gaussian, is passed through a non-linear function *g(x).* The information in the resulting density, *p(x)* depends on matching the mean and variance of **x** to the threshold, $\mathbf{w_0}$, and slope, **w**, of *g(x)* (Nicol Schraudolph, personal communication). (*right*) *p(y)* is plotted for different values of the weight **w**. The optimal weight, $\mathbf{w_{opt}}$ transmits most information (from Bell & Sejnowski, 1995).

The match cannot be perfect, but it does approach the maximum entropy p.d.f. for the unit distribution by maximizing the expected log slope, $E[\log|g'(\mathbf{Wx})|]$.

The generalization of this idea to **N** dimensions leads to maximizing the expected log determinant of the absolute value of the Jacobian matrix $|[\partial g_i(u_i)/\ \partial x_j]_{ij}|$. This optimization attempts to map the input vectors uniformly into the unit **N**-cube (assuming that the **g**-functions are still 0-1 bounded). Intuitively, if the outputs are spread evenly (like molecules of a gas) throughout their (N-cube) range, then learning the value of a data point on one axis gives no information about its values on the other axes and maximum independence has been achieved. Bell and Sejnowski (1995) showed that the stochastic gradient descent algorithm that maximizes **H(g(u))** is:

$$\Delta\mathbf{W} \propto \mathbf{W}^{-T} + \mathbf{f}(\mathbf{u})\mathbf{x}^T \tag{8}$$

where $-T$ denotes inverse transpose, and the vector-function, **f**, has elements:

$$f_i(u_i) = \frac{\partial}{\partial u_i} \ln g'_i(u_i) \tag{9}$$

When $g'_i(u_i) = p_i(u_i)$ for all i then, according to (6), the ICA algorithm is exact. Unfortunately, this leaves a difficulty. Either one has to estimate the functions **g** during training, or one needs to assume that the final term in (6) does not interfere with Infomax performing ICA. We have empirically observed a systematic robustness to mis-estimation of the prior, $\hat{p}_i(u_i \mid \mathbf{W}, \mathbf{g}) = |g'_i(u_i)|$. Although unproven, this Robustness Conjecture can be stated: Any super-Gaussian prior will suffice to extract super-Gaussian independent components. Any sub-Gaussian prior will suffice to extract sub-Gaussian independent components. This conjecture also leads to the generally successful 'extended ICA' algorithms (Girolami, 1998; Lee, Girolami & Sejnowsk, 1999) that switch the component priors, $\hat{p}_i(u_i)$, between super- and sub-Gaussian functions. In practice, as the robustness principle suggests, this switching may be all the estimation needed to obtain a correct solution. The same insight underlies 'negentropy' approaches to ICA that maximize the distance of the $p_i(u_i)$ from Gaussian, described by Hyvaerinen (1999) and Lee, Girolami and Sejnowsk (1999).

For most natural data (images, sounds etc), the independent component p.d.f.'s are all super-Gaussian, so many good results have been achieved using 'logistic ICA,' in which the super-Gaussian prior is the slope, $g'_i(u_i)$, of the common logistic sigmoid function (8) so often used in neural networks. For this choice of g, the function f in (8) evaluates simply to $f(u) = 1-2g(u)$.

An additional and important feature was added to the Infomax ICA algorithm by Amari (1998), who observed that a simpler learning rule, with much faster and more stable convergence, could be obtained by multiplying the Infomax gradient of (8) by $\mathbf{W}^T\mathbf{W}$, obtaining:

$$\Delta\mathbf{W}_{NatGrad} = (\Delta\mathbf{W})\mathbf{W}^T\mathbf{W} \propto \left(\mathbf{I} + \mathbf{f}(\mathbf{u})\mathbf{u}^T\right)\mathbf{W} \tag{10}$$

Since $\mathbf{W}^T\mathbf{W}$, which scales the gradient, is positive-definite, it does not change the minima and maxima of the optimization. Its optimality has been explained using information geometry (Amari, 1998) and equivariance – the gradient vector local to **W** is normalized to behave as if it were close to **I** (see Haykin, 2000). Both interpretations reflect the fact that the parameter space of **W** is not truly Euclidean, since its axes are entries of a matrix. Equation (10) is clearly a nonlinear decorrelation rule, stabilizing when $\langle -\mathbf{f}(\mathbf{u})\mathbf{u}^T \rangle = \mathbf{I}$. (The minus sign is required because the **f** functions are typically decreasing). The Taylor series expansion of the **f** functions provides information about higher-order correlations necessary to perform ICA.

4.5.4.3 State of the Art

Several studies have been performed to demonstrate the power of the ICA algorithm to analyse biomedical data. Biomedical signals are a rich source of information about physiological processes, but they are often contaminated with artifacts or noise and are typically mixtures of unknown combinations of sources summing differently at each of the sensors. For example, the electroencephalographic (EEG) data is a non-invasive measure of brain electrical activity recorded as changes in potential difference between points on the human scalp. Because of volume conduction through cerebrospinal fluid, skull and scalp, EEG data collected from any point on the scalp may include activity from multiple processes occurring within a large brain volume. This has made it difficult to relate EEG measurements to underlying brain processes or to localize the sources of the EEG signals. Makeig and co-workers (Makeig & Jung, 1996; Makeig, Jung, Ghahremani, Bell & Sejnowski, 1997; Makeig, et al., 2002) first applied the original infomax algorithm to EEG and event-related potential (ERP) data showing that the algorithm can extract

EEG activations and isolate artifacts. Jung and colleagues (Jung, et al., 1998, 2000a, 2000b, Jung, et al., 2001a, Jung, et al., 2001b) and Makeig et al. (2002) showed that the extended infomax algorithm (Lee, Girolami, & Sejnowski, 1999) allows us to: (1) remove pervasive artifacts of all types from single-trial EEG records, making possible analysis of highly contaminated EEG records from clinical populations; (2) identify and segregate stimulus- and response-locked event-related activity in single-trail EEG epochs following stimulus presentation; (3) separate spatially-overlapping EEG activities over the entire scalp and frequency band that may show a variety of distinct relationships to task events, rather than focusing on activity at single frequencies in single scalp channels or channel pairs. (4) investigate the interaction between ERPs and ongoing EEG.

McKeown, et al. (1998a, 1998b) demonstrated for the first time, that ICA can also be used to analyze hemodynamic signals from the brain recorded using functional magnetic resonance imaging (fMRI). The FMRI technique is a non-invasive technique used to localize dynamic brain processes in intact living brains (Kwong et al., 1992). It is based on the magnetic susceptibilities of oxygenated hemoglobin (HbO2) and deoxygenated hemoglobin (HbR) and is used to track blood-flow-related phenomena accompanying or following neuronal activations. The most commonly used fMRI signal is the blood-oxygen-level-dependent (BOLD) contrast (Ogawa et al., 1992). ICA, applied to fMRI data, has proven to be a powerful method for detecting task-related activations, including unanticipated activations (McKewon et al., 1998a; 1998b) that could not be detected by standard hypothesis-driven analyses. This may expand the types of fMRI experiments that can be performed and meaningfully interpreted.

Other interesting applications of ICA are to the electrocorticogram (EcoG) – direct measurements of electrical activity from the surface of the cortex, and to optical recordings of electrical activity from the surface of the cortex using voltage-sensitive dyes (Schiessbl, et al., 2000). ICA has also proven effective at analyzing single-unit activity from the cerebral cortex (Laubach, Shuler, & Nicolelis, 1999; Laubach, Wessberg & Nicolelis, 2000) and in separating neurons in optical recordings from invertebrate ganglia (Brown, Yamada, & Sejnowski, 2001). Early clinical research applications of ICA include the analysis of EEG recordings during epileptic seizures (McKeown, Humphries, Iragui & Sejnowski, 1999).

In addition to the brain signals that were the focus of this paper, signals from other organs, including the heart (Jung et al., 2000) and endocrine system (Prank, Borger, von zur Muhlen, Brabant & Schofl, 1999) have similar problems with artifacts that could also benefit from ICA. Bartlett and Sejnowski, (1997), Bartlett, et al., (2000) and Gray, Movellan and Sejnowski (1997) demonstrated the successful use of the ICA filters as features in face recognition and lip reading tasks, respectively.

ICA is a fairly new but powerful analysis capability. Results often reveal novel data structures, which provoke a diversity of theoretical questions. ICA software can be downloaded from several laboratory websites and utilized for a variety of complex problems.

4.5.4.4 References

Amari, S. (1998). Natural Gradient Works Efficiently in Learning. *Neural Computation*, *10*, 251-76.

Bartlett, M.S., Donato, G.L., Movellan, J.R., Hager, J.C., Ekman, P., & Sejnowski, T.J. (2000). Image Representations for Facial Expression Coding. In S. Solla, T. Leen, & K. Mueller, (Eds.), *Advances in Neural Information Processing Systems 12*, Cambridge, MA: MIT Press, (pp. 886-892).

Bartlett, M.S., & Sejnowski, T.J. (1997). Viewpoint Invariant Face Recognition using Independent Component Analysis and Attractor Networks. In: M. Mozer, M. Jordan, & T. Petsche (Eds.), *Advances in Neural Information Processing Systems 9*, MIT Press, Cambridge, MA. (pp. 817-823).

Bell, A.J., & Sejnowski, T.J. (1995). An Information Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computing*, *7*, 1129-1159.

Brown, G.D., Yamada, S., & Sejnowski, T.J. (2001). Independent Component Analysis (ICA) at the Neural Cocktail Party. *Trends in Neuroscience*, *24*, 54-63.

Cardoso, J.F., & Laheld, B.H. (1996). Equivariant Adaptive Source Separation. *IEEE Transactions on Signal Processing*, *44*, 3017-3030.

Girolami, M. (1998). An Alternative Perspective on Adaptive Independent Component Analysis Algorithm. *Neural Computation*, *10*, 2103-2114.

Comon, P. (1994). Independent Component Analysis – A New Concept? *Signal Processing*, *36*, 287–314.

Gray, M.S., Movellan, J.R., & Sejnowski, T.J. (1997). A Comparison of Local versus Global Image Decompositions for Visual Speechreading. *Proceedings of the 4th Joint Symposium on Neural Computation* (pp. 92-98).

Haykin, S. (2000). *Unsupervised Adaptive Filtering*. New York: Prentice Hall.

Herault, J., & Jutten, C. (1986). Space or Time Adaptive Signal Processing by Neural Networks Model. International Conference on Neural Networks for Computing, Snowbird (Utah, USA). *AIP Conference Proceedings*, J. Denker (Ed.), (pp. 206-211).

Hyvaerinen, A. (1999). Survey on Independent Component Analysis. *Neural Computing Surveys*, *2*, 94-128.

Jung, T.-P., Humphries, C., Lee, T.W., Makeig, S., McKeown, M.J., Iragui, V., & Sejnowski, T.J. (1998). Extended ICA Removes Artifacts from Electroencephalographic Recordings. *Advances in Neural Information Processing Systems 10*, 894-900.

Jung, T.-P., Makeig, S., Lee, T.-W., McKeown, M.J., Brown, G., Bell, A.J., & Sejnowski, T.J. (2000a). Independent Component Analysis of Biomedical Signals, *The 2nd International Workshop on Independent Component Analysis and Signal Separation* (pp. 633-644), Helsinki, Finland.

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M., Iragui, V., & Sejnowski, T.J. (2000b). Removing Electroencephalographic Artifacts by Blind Source Separation. *Psychophysiology*, *37*, 167-178.

Jung, T.-P, Makeig, S., McKeown, M.J., Bell, A.J., Lee, T.-W., & Sejnowski, T.J. (2001a). Imaging Brain Dynamics using Independent Component Analysis. *Proceedings of the IEEE*, *89*, 1107-1122.

Jung, T.-P., Makeig, S., Westerfield, W., Townsend, J., Courchesne, E., & Sejnowski, T.J. (2001b). Analysis and Visualization of Single-Trial Event-Related Potentials. *Human Brain Mapping*, *14*, 166-185.

Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E., Cohen, M.S., & Turner, R. (1992). Dynamic Magnetic Resonance Imaging of Human Brain Activity during Primary Sensory Stimulation. *Proceedings of the National Academy of Sciences USA*, *89*, 5675-5679.

Laubach, M., Shuler, M., & Nicolelis, M.A.L. (1999). Independent Component Analyses for Quantifying Neuronal Ensemble Interactions. *Journal of Neuroscience Methods*, *94*, 141-154.

Laubach, M., Wessberg, J., & Nicolelis, M.A.L. (2000). Cortical Ensemble Activity Increasingly Predicts Behavioral Outcomes during Learning of a Motor Task. *Nature*, *405*, 567-571.

Lee, T.-W. Girolami, M., & Sejnowski, T.J. (1999). Independent Component Analysis using an Extended Infomax Algorithm for Mixed Sub-Gaussian and Super-Gaussian Sources. *Neural Computing, 11*, 409-433.

Makeig, S., & Jung, T.-P. (1996). Tonic, Phasic, and Transient EEG Correlates of Auditory Awareness during Drowsiness. *Cognitive Brain Research, 4*, 15-25.

Makeig, S., Jung, T.-P., Ghahremani, D., Bell, A.J., & Sejnowski, T.J. (1997). Blind Separation of Event-Related Brain Responses into Independent Components. *Proceedings of the National Academy of Sciences USA*, *94*, 10979-10984.

Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E., & Sejnowski, T.J. (2002). Dynamic Brain Sources of Visual Evoked Responses. *Science*, *295*, 690-694.

McKeown, M., Jung, T.-P., Makeig, S., Brown, G., Kindermann, S., Lee, T.-W., & Sejnowski, T.J. (1998a). Spatially Independent Activity Patterns in Functional Magnetic Resonance Imaging Data during the Stroop Color-Naming Task. *Proceedings of the National Academy of Sciences*, *95*, 3-10.

McKeown, M., Makeig, S., Brown, G.G., Jung, T.-P., Kindermann, S.S., Bell, A.J., & Sejnowski, T.J. (1998b). Analysis of fMRI Data by Blind Separation into Independent Spatial Components. *Human Brain Mapping*, *6*, 160-188.

McKeown, M., Humphries, C., Iragui, V., & Sejnowski, T.J. (1999). Spatially Fixed Patterns Account for the Spike and Wave Features in Absence Seizures. *Brain Topography*, *12*, 107-116.

Ogawa, S., Tank, D., Menon, R., Ellermann, J., Kim, S., Merkle, H., & Ugurbil, K. (1992). Intrinsic Signal Changes Accompanying Sensory Stimulation: Functional Brain Mapping with Magnetic Resonance Imaging. *Proceedings of the National Academy of Sciences USA*, *89*, 51-55.

Pham, D.T., Garat, P., & Jutten, C. (1992). Separation of a Mixture of Independent Sources through a Maximum Likelihood Approach. In J. Vandewalle, R.L. Boite, M. Moonen, & A. Oosterlinck (Eds.), *Signal Processing VI, Proceedings of EUSIPCO '92* (pp. 771-774).

Prank, K., Borger, J., von zur Muhlen, A., Brabant, G., & Schofl, C. (1999). Independent Component Analysis of Intracellular Calcium Spike Data. *Advances in Neural and Information Processing Systems*, *11*, 931-937.

Schiessbl, I., Stetter, M., Mayhew, J.E., McLoughlin, N., Lund, J.S., & Obermayer, K. (2000). Blind Signal Separation from Optical Imaging Recordings with Extended Spatial Decorrelation. *IEEE Transactions on Biomedical Engineering*, *47*, 573-577.

Chapter 5 – PROSPECTIVE OUTLOOK

5.1 TECHNOLOGY INNOVATIONS

One of the obstacles to widespread application of psychophysiological methods is the problem of electrode application. Many operators do not like to wear electrodes during their duty hours. Several methods of non-contact sensing of physiological activity and remote sensing are available or undergoing development. Various methods of monitoring heart rate using ballistocardiography and wide-band technology are being attempted in operational situations. If successfully tested in operational environments, heart rate will be measured without applying any sensors to the operators. Remote sensing of eye point of regard, eye blinks, and pupil diameter have already been discussed in this report. These measures are collected without applying sensors to the operators and have been tested in workstations and while driving vehicles. Rapid application electrodes have been developed which require minimal skin preparation. Dry electrodes that can be used for recording cardiac and brain electrical activity have been demonstrated and are under development. Other devices such as NIRS, discussed earlier, record brain metabolic activity. These sensors are typically held in place with a band or cap and do not intrude on the operator's task.

The continuing validation of Moore's Law, which predicts the doubling of computer processing speed and memory capacity every 18 months, bodes well for the implementation of OFS in operational systems. Typically, the data sampling rates required for OFS assessment are low, but the number of channels sampled and the hours of duty time yield large amounts of data. As mentioned above in the discussions of several of the measures, signal processing must often be applied to the raw data to detect and correct artifacts, to condition the data for further analysis, and to perform the OFS assessment. The automatic interpretation of the processed data must be accomplished within the context of the current mission requirements. Currently available high-speed PC processors, large computer memories, and storage devices with vast storage capacity make possible processing that was not achievable just a few years ago. Telemetry systems can be used so that operators are not tethered to the recording equipment. The use of wireless LANs permits collection of data simultaneously from a number of operators and allows assessment of the state of an entire crew in a multiple-person system. The continued development of hardware and software capabilities will make possible even more powerful OFS assessment systems in the future.

5.2 ADAPTIVE AIDING

Current systems operated by humans are very complex and future systems will no doubt increase in complexity. These systems are able to overwhelm the capacity of the human operator with the amount of information presented, the complexity of the decisions to be made and the speed with which these decisions must be made. However, along with the added complexity is the ability of the system to carry out some of the tasks that were traditionally the job of the human operator. Adaptive aiding is the term used to describe the situation where the system assumes functions that the human normally is in charge of. The adaptive aiding is implemented when it is determined that the operator requires help. This could be in the case of cognitive overload where the cognitive capabilities of the operator have been exceeded by the task demands. In order for the adaptive aiding to assist and not hinder the operator it must be presented only when needed. The correct determination of the OFS is crucial to this. The aiding must only be presented when required. If presented at inappropriate times the aiding may add to the cognitive load of the operator and make the situation worse. In order for adaptive aiding to assist the operator to lower errors, accurate assessment of OFS is critical. If the operator's state can be accurately assessed then the aiding will have the desired beneficial effects. Laboratory studies have demonstrated the utility of psychophysiological measures to assess the operator's state with a high degree of accuracy.

Furthermore, the benefit of adaptive aiding has been shown in the laboratory setting Successful transition of these techniques to real-world systems will both improve system performance and reduce errors.

5.3 PREDICTION IMPROVEMENT

The ability to predict future functional state from current and past state is a key goal of any operational assessment system. Given appropriate system capability it will be possible to make modifications in workload, environmental stress, or clothing and equipment. Advances in model development, fuzzy logic, neural nets, and statistical techniques combined with increasing processing power will allow for the implementation of real-time multiple-trend predictor techniques using physiological and operational data. Both short-term (on the order of seconds and minutes) and long-term (hours and days) prediction will be required depending on operational requirements.

The success of adaptive aiding using OFS will depend upon highly accurate prediction of OFS. In fact, the implementation of adaptive aiding will only work if the accuracy of the assessment is very high. Otherwise, the aiding will not be presented when needed or, on the other hand, will be presented when it is not needed. The ability to accurately predict further into the future will also help with the success of adaptive aiding. The farther in the future that performance breakdown can be predicted, the sooner intervention can be made to maintain optimal system performance.

5.4 READINESS TO PERFORM

A different approach can be used to develop performance indices using physiological parameters. The accuracy of such assessment can be high, but it is difficult to accomplish in operational environments. However, this approach uses pre-shift or pre-mission assessment. On the basis of this assessment one is able to make a prediction of the operator’s fitness (readiness)-for-duty. This is done on the basis of test performance on a standard task that yields indirect indices of the operator’s functional state to determine fitness-for-duty. This approach has been successfully applied to operators in the power industry, transportation, and the training of flying personnel. This approach should be tested in other fields of application.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	RTO-TR-HFM-104 AC/323(HFM-104)TP/48	ISBN 92-837-1111-4	UNCLASSIFIED/ UNLIMITED

5. Originator Research and Technology Organisation
North Atlantic Treaty Organisation
BP 25, F-92201 Neuilly-sur-Seine Cedex, France

6. Title
Operator Functional State Assessment

7. Presented at/Sponsored by
This Technical Report has been prepared by the RTO Human Factors and Medicine Panel (HFM) Task Group HFM-056/TG-008. The material in this publication also supported a Lecture Series under the sponsorship of the Human Factors and Medicine Panel (HFM) presented on 8-9 September 2003 in Kiev, Ukraine; 11-12 September 2003 in Brussels, Belgium; and 2-3 October 2003 in San Diego, USA.

8. Author(s)/Editor(s)	9. Date
Multiple	February 2004

10. Author's/Editor's Address	11. Pages
Multiple	220

12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.

13. Keywords/Descriptors

Acceleration stresses (physiology)
Aerospace environment
Anthropometry
Biomedical measurement
Environments
Human factors engineering
Job analysis
Models
Noise (sound)
Operational effectiveness
Operators (personnel)
Performance evaluation
Reviews
Risk
Scenarios
Situational awareness
Stress (physiology)
Systems analysis
Systems engineering
Task analysis
Vibration
Work measurement
Workloads
Workplace layout

14. Abstract

The human operator is a crucial component of complex modern systems. The complexity of these systems, the rapid tempo of contemporary military operations and reduced staffing all contribute to the high cognitive demands experienced by military personnel. The rate of information flow, the number of decisions, and actions that must be carried out can exceed the cognitive capacity of the operator. This can result in disastrous consequences. Other system components are continuously scrutinized so that problems can be identified and rectified. The functional state of the human operator is not monitored and remediation is not implemented. Proper assessment of the functional state of the operator and appropriate intervention will result in improved mission effectiveness.

The goal of this report is to assemble the pertinent information concerning the factors that produce suboptimal performance in human operators. Numerous methods are available to detect the presence of these factors. This report provides a comprehensive survey of the risk factors that impact human performance and the assessment methods for measuring these effects. The risk factors include environmental features such as noise, acceleration and thermal stress. States within the individual operator can interfere with optimal performance and include illness, sleep loss and disruption of circadian rhythms. Task characteristics include the cognitive and physical demands of the task. Theoretical concerns are presented as a framework for the risk factors that reduce the functioning of human operators. Methods for detecting impaired operator functional state are presented and include physiological, performance, and subjective assessment procedures. The rationale for each measure is presented along with the technological required to make the measurements.

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DIFFUSION DES PUBLICATIONS
RTO NON CLASSIFIEES

Les publications de l'AGARD et de la RTO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus soit à titre personnel, soit au nom de votre organisation, sur la liste d'envoi.

Les publications de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivi du numéro de série. Des informations analogues, telles que le titre est la date de publication sont souhaitables.

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la RTO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (www.rta.nato.int) et vous abonner à ce service.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Streitkräfteamt / Abteilung III
Fachinformationszentrum der
Bundeswehr (FIZBw)
Friedrich-Ebert-Allee 34, D-53113 Bonn

BELGIQUE
Etat-Major de la Défense
Département d'Etat-Major Stratégie
ACOS-STRAT – Coord. RTO
Quartier Reine Elisabeth
Rue d'Evère, B-1140 Bruxelles

CANADA
DSIGRD2
Bibliothécaire des ressources du savoir
R et D pour la défense Canada
Ministère de la Défense nationale
305, rue Rideau, 9e étage
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1, P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

ETATS-UNIS
NASA Center for AeroSpace
Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GRECE (Correspondant)
Defence Industry & Research
General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HONGRIE
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Centro di Documentazione
Tecnico-Scientifica della Difesa
Via XX Settembre 123
00187 Roma

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25, NO-2007 Kjeller

PAYS-BAS
Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

POLOGNE
Armament Policy Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

REPUBLIQUE TCHEQUE
DIC Czech Republic-NATO RTO
VTÚL a PVO Praha
Mladoboleslavská ul.
Praha 9, 197 06
Česká republika

ROYAUME-UNI
Dstl Knowledge Services
Information Centre, Building 247
Dstl Porton Down
Salisbury
Wiltshire SP4 0JQ

TURQUIE
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320
ETATS-UNIS

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
ROYAUME-UNI

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Acquisitions, Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants :

Scientific and Technical Aerospace Reports (STAR)
STAR peut être consulté en ligne au localisateur de ressources uniformes (URL) suivant:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR est édité par CASI dans le cadre du programme NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
ETATS-UNIS

Government Reports Announcements & Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 2216
ETATS-UNIS
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DISTRIBUTION OF UNCLASSIFIED RTO PUBLICATIONS

AGARD & RTO publications are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO reports, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your Organisation) in their distribution.

RTO and AGARD reports may also be purchased from the Sales Agencies listed below.

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number. Collateral information such as title and publication date is desirable.

If you wish to receive electronic notification of RTO reports as they are published, please visit our website (www.rta.nato.int) from where you can register for this service.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Etat-Major de la Défense
Département d'Etat-Major Stratégie
ACOS-STRAT – Coord. RTO
Quartier Reine Elisabeth
Rue d'Evère
B-1140 Bruxelles

CANADA
DRDKIM2
Knowledge Resources Librarian
Defence R&D Canada
Department of National Defence
305 Rideau Street
9th Floor
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC
DIC Czech Republic-NATO RTO
VTÚL a PVO Praha
Mladoboleslavská ul.
Praha 9, 197 06
Česká republika

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72
92322 Châtillon Cedex

GERMANY
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Friedrich-Ebert-Allee 34
D-53113 Bonn

GREECE (Point of Contact)
Defence Industry & Research General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HUNGARY
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ICELAND
Director of Aviation
c/o Flugrad, Reykjavik

ITALY
Centro di Documentazione Tecnico-Scientifica della Difesa
Via XX Settembre 123
00187 Roma

LUXEMBOURG
See Belgium

NETHERLANDS
Royal Netherlands Military Academy Library
P.O. Box 90.002
4800 PA Breda

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25, NO-2007 Kjeller

POLAND
Armament Policy Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide, P-2720 Amadora

SPAIN
SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

TURKEY
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

UNITED KINGDOM
Dstl Knowledge Services
Information Centre, Building 247
Dstl Porton Down
Salisbury, Wiltshire SP4 0JQ

UNITED STATES
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320

SALES AGENCIES

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320
UNITED STATES

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
UNITED KINGDOM

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
STAR is available on-line at the following uniform resource locator:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR is published by CASI for the NASA Scientific and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
UNITED STATES

Government Reports Announcements & Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 2216
UNITED STATES
(also available online in the NTIS Bibliographic Database or on CD-ROM)

ISBN 92-837-1111-4