AFIT/DS/ENG/05-01

Operator State Estimation for Adaptive Aiding in Uninhabited Combat Air Vehicles

Dissertation

Christopher A. Russell

AFIT/DS/ENG/05-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.

AFIT/DS/ENG/05-01

OPERATOR STATE ESTIMATION FOR ADAPTIVE AIDING IN UNINHABITED COMBAT AIR VEHICLES

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Christopher A. Russell, BSEE, MSEE

September 2005

AFIT/DS/ENG/05-01

OPERATOR STATE ESTIMATION FOR ADAPTIVE AIDING IN UNINHABITED COMBAT AIR VEHICLES

Christopher A. Russell, BSEE, MSEE

Approved: Date:

_____/SIGNED/_____ _____
Steven C. Gustafson (Chairman)

_____/SIGNED/_____ _____
Peter S. Maybeck (Dean's Representative)

_____/SIGNED/_____ _____
Mateen M. Rizki (Member)

_____/SIGNED/_____ _____
Timothy S. Webb (Member)

_____/SIGNED/_____ _____
Glenn F. Wilson (Member/Sponsor)

Accepted:

_____ _____
Robert A. Calico, Jr. Date
Dean, Graduate School of Engineering and Management

Abstract

This research demonstrated the first closed-loop implementation of adaptive automation using operator functional state in an operationally relevant environment. In the Uninhabited Combat Air Vehicle (UCAV) environment, operators can become cognitively overloaded and their performance may decrease during mission critical events. Additionally, pervasive automation could degrade UCAV operator situation awareness and capability to react appropriately to unusual events. The critical question, therefore, was if automation could be used adaptively to allow the operator to deal effectively with high workload situations without excessive disengagement from the task. Researchers have attempted to use operator functional state to guide adaptive aiding but never accomplished it in an operationally relevant task environment. This research, however, demonstrates an unprecedented closed-loop system, one that adaptively aids UCAV operators based on their cognitive functional state.

The operator functional state was determined by integrating and assessing multiple psychophysiological measures using an operator state classification system. That system was then used to change the environment and allow the operator to improve performance. A series of experiments were conducted to 1) determine the best classifiers for estimating operator functional state, 2) determine if physiological measures can be used to develop multiple cognitive models based on information processing demands and task type, 3) determine the salient psychophysiological measures in operator functional state, and 4) demonstrate the benefits of intelligent adaptive aiding using operator functional state.

Single-task experiments, representing subtasks of the suppression of enemy air defenses (SEAD) mission, were conducted for six operators. One subtask required the operator to monitor vehicle health status and initiate corrections or repairs periodically. The second subtask required the operator to determine and select targets in synthetic

aperture radar (SAR) images. These experiments were used for classifier comparisons, feature saliency analysis, and cognitive model development.

Next, three types of classification algorithms were compared, including artificial neural networks, discriminant analysis, and support vector machines. In general, nonlinear classifiers or linear classifiers implemented after a nonlinear transformation performed best. That is, the multilayer perceptron classifier with backpropagation training outperformed linear and quadratic discriminant analysis, logistic regression, and linear and radial basis function support vector machines. The multilayer perceptron outperformed the other classifiers in 58 to 80% of the comparisons.

Several models were developed using multilayer perceptron classifiers to determine the utility of applying the same psychophysiological measures as inputs and to identify multiple cognitive gauges. Gauges identifying levels of cognitive difficulty in spatial working memory, verbal working memory, executive function, spatial versus verbal working memory, global workload, vehicle health task, and SEAD tasks were developed. Classification accuracy for all cognitive gauges ranged from 59 to 91%.

To determine the effects of adaptive aiding in a complex operational environment, experiments were conducted with operators who performed the SEAD missions and vehicle health tasks in a UCAV simulator. Adaptive aiding was implemented using operator state estimation as a control input that adapts the system when the operator is cognitively loaded. Aiding the operator actually improved performance and increased mission effectiveness by 67% in that missed weapons release, which indicates mission failure, is reduced by this percentage. It was found that the operators must be aided at appropriate times; operators aided at random times had the same performance as unaided operators.

Acknowledgments

I would like to thank my committee, Dr. Gustafson, Dr. Rizki, Maj. Webb, and Dr. Wilson for their guidance in the completion of this research. I would like to thank Dr. Glenn Wilson especially for sponsoring this research and for championing the idea of operator state estimation for many years. I would also like to thank all the folks in the lab and in my branch for supporting me in this endeavor.

A special thanks to my nephews who are five and three and always make me smile. Thanks, for your help editing this paper using most of the colors in a Crayola 64 pack and for helping me type my edits. This next sentence is for you. Ccccccclllllllllllllllhhhhhhh.

I thank my lovely bride for her never ending love and support. I would like to thank my Mom and Dad and my wife's mom and dad for their support and confidence in me. Thanks, Mom, for always saying to me, "Look it up."

Table of Contents

Page

List of Figures

List of Tables

Table Page

List of Abbreviations

AFIT	Air Force Institute of Technology
AFRL	Air Force Research Laboratory
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
BCI	Brain Computer Interface
DA	Discriminant Analysis
DARPA	Defense Advanced Research Programs Agency
DMPI	Designated Mean Point of Impact
ECG	Electrocardiogram
EOG	Electo-oculogram
EEG	Electroencephalography
HSD	Honestly Significant Difference
IBI	Interbeat Interval
IBLI	Interblink Interval
ICA	Independent Component Analysis
LDA	Linear Discriminant Analysis
OVI	Operator Vehicle Interface
PCA	Principal Component Analysis
QDA	Quadratic Discriminant Analysis
REM	Rapid Eye Movement
SNR	Signal-to-Noise Ratio
SEAD	Suppression of Enemy Air Defense

SAR	Synthetic Aperture Radar
SVM	Support Vector Machine
SWDA	Stepwise Discriminant Analysis
UAV	Uninhabited Air Vehicle
UCAV	Uninhabited Combat Air Vehicle
USAF	United States Air Force
VHT	Vehicle Health Task

OPERATOR STATE ESTIMATION FOR ADAPTIVE AIDING IN UNINHABITED COMBAT AIR VEHICLES

I. Introduction

1.1 Research Accomplishments

This dissertation presents the first implementation of closed-loop real-time adaptive aiding using operator functional state in an operationally relevant environment: the Uninhabited Combat Air Vehicle (UCAV). Improvements in operator performance on mission critical measures, such as the number of targets hit, demonstrated the utility of adaptive aiding. Meeting the overall objective of this research required a robust operator state classification, one used in intelligent adaptive aiding to improve human-machine performance in military systems. Additionally, psychophysiological measures, both new nonstandard and traditional were developed, identified, extracted, and integrated in the classification system.

The focus of a $70M DARPA Augmented Cognition Program and a major thrust of the Air Force Research Laboratory (AFRL) program on future human-machine collaborative systems, this research significantly extended previous AFIT research and made following significant contributions:

- It established the first example of adaptive aiding using operator functional state in an operationally relevant environment. Adaptive aiding was implemented in a real-time closed loop system using operator functional state in a UCAV simulator.
- It demonstrated significant improvement in mission effectiveness using adaptive aiding. The implementation of adaptive aiding reduced the

occurrence of missed weapons release waypoints from 25% in the trials without adaptive aiding to 8% in the trials with adaptive aiding, which was a 67% improvement in mission effectiveness.

- It represented the first exploration of multiple cognitive model development defined by information processing demands and task type. Models were developed for spatial working memory, verbal working memory, executive function, global workload, spatial versus verbal working memory, vehicle health task identification, and operator vehicle interface task identification.
- It demonstrated the identification, integration, and extraction of multiple psychophysiological measures into a cognitive operator functional state model. Features were derived from electroencephalography (EEG), electrocardiography (ECG), electro-oculagraphy (EOG), electromyography (EMG), and electrodermal signals and integrated into an operator functional state model.
- It made a direct comparison of multiple types of pattern classification methods using 'real-world' psychophysiological data. Classification algorithms based on artificial neural networks, support vector machines, and discriminant analysis were compared directly to determine their utility in classifying operator functional state.

This research resulted in several publications and presentations:

Russell, Chris A. "Statistical and Mathematical Tools: Artificial Neural Networks" in Operator Functional State Assessment: Optimizing Systems Performance, *NATO RTO Technical Report*, Kiev, Ukraine, Brussels, Belgium, and San Diego, USA, December 2003.

Wilson, Glenn F. and Chris A. Russell. "Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks," *Human Factors*, Winter 2003.

Russell, Chris A. Team State Classification Methods, *Augmented Cognition PI Meeting*, Orlando, FL., 5–8 January 2004.

Russell, Chris A. Operator State Estimation Workshop, Invited Speaker, *Augmented Cognition PI Meeting*, Orlando, FL, 5–8 January 2004.

Russell, Chris A. Operator State Estimation, Invited Lecturer, Wright State University, *EGR 861 PhD Seminar*, February 20, 2004.

Wilson, Glenn F. and Chris A. Russell. "Psychophysiologically Determined Adaptive Aiding in a Simulated UCAV Task," *Human Performance, Situation Awareness and Automation Technology Conference*, Daytona Beach, FL., 22-25 March 2004.

Wilson, Glenn F. and Chris A. Russell. "Psychophysiologically Determined Classification of Cognitive Activity", *Human Factors Conference*, November 2004.

Russell, Chris A. Lecturer, *Human Interfaces Course*, AFIT, 28 February 2005.

Russell, Chris A., Glenn F. Wilson, Mateen M. Rizki, Timothy S. Webb, and Steven C. Gustafson. "Comparing Classifiers for Real Time Estimation of Cognitive Workload," *Human Computer Interface Conference*, Las Vegas NV, 25-27 July 2005.

1.2 Overview

The complexity of advanced military systems is increasing and has generated interest in the interface between the human operator and complex systems. In some situations, system complexity can overwhelm the human operator. The interface is usually inflexible, or at the very least, difficult to manipulate in real time. The operator, unaware that trouble exists, may shed less demanding tasks to complete the immediate task. The operator may become "overloaded" resulting in decreased operator performance, decreased situational awareness, or mission failure.

1.2.1 The Nature of Adaptive Aiding

The traditional method of increasing operator performance and reducing operator workload has been to make static improvements in the interface between the machine and the human operator. Dynamically modifying the interface based on operator need could be an alternate approach. By measuring operator functional state or operator ability to accomplish current tasks, the system interface could be adapted or modified to aid the operator in performing the assigned task. As such, adaptive automation could improve operator performance and reduce operator workload by adapting the interface "on demand" based on operator needs and functional state.

The implementation of adaptive aiding using operator functional state required developing and integrating several areas of research. The components of operator functional state assessment were defined and modeled. Pattern classification algorithms were evaluated to determine the appropriate choice for use in the classification of operator functional state. Appropriate techniques for adaptive automation were determined for improved operator performance and reduced operator cognitive workload. Finally, these areas were integrated and evaluated in an operational environment. This research addressed all these issues to some degree, and a brief overview is provided in the remainder of this section.

Operator state assessment consists of four major components: psycho-physiological assessment (cognitive workload), operator performance assessment, situation awareness assessment, and momentary mission requirements (Gaillard and Kramer, 2000; Wilson, 2003). Models for each component are necessary for accurate operator state assessment and, in turn, intelligent adaptive aiding. Aiding may not be

required if any of these components peak or trough individually. Also, application of intelligent adaptive aiding is not required continuously. Rather, appropriate aiding is needed when the operator cannot perform the tasks required or when a decrease in task load is necessary for completing the mission. The primary motivation of this research, however, is to provide robust real-time human cognitive state estimation and apply such estimation for adaptive decision aiding in complex task environments.

Estimation of operator state has numerous applications in the fields of human factors engineering, training, testing, and evaluation. For instance, Uninhabited Air Vehicle (UAV) and UCAV operators may experience performance degradation during mission segments with high cognitive load. An understanding of operator workload could aid in the development of human-computer interfaces by providing metrics for operator state. In addition, accurate and reliable assessment of operator state is key to successful implementation of adaptive automation, design evaluation, and operational test and evaluation. Although, real-time operator functional state estimation has been historically limited by the processing capabilities of computers, the advent of increased processing power now permits complex inference models to classify operator functional state in real time.

1.2.2 Models for Adaptive Aiding

Classical statistical inference is based on three fundamental assumptions (Casella and Berger, 2002; Scharf, 1991). First, data can be modeled by a set of linear functions. Unfortunately, real-world problems are often high-dimensional, and the underlying mapping is usually not very smooth. Under these conditions linear paradigms need a large number of terms. Also, high dimensionality of the input space implies a large

number of independent variables, which leads to "the curse of dimensionality" (Gershenfeld, 1999). Second, the underlying joint probability density is assumed to be Gaussian (i.e., normal), which may not be the case for real data; the data may be far from normally distributed. Finally, due to the second assumption, the usual induction paradigm for parameter estimation is the maximum likelihood method; it reduces to the minimization of a sum of squared error cost function in most engineering problems but can be inappropriate.

An artificial neural network (ANN) can in principle address all these concerns. ANNs have advantages that make them potential classifiers of operator cognitive state. Because of the inherent nonlinearity and the complex interactions among the features of cognitive activity during dynamic multiple task situations, accurate workload classification is difficult. Further, the relationships between physiological variables and performance are complex, and highly dynamic tasks are not well understood; therefore, the relevant features for cognitive workload classification in these highly dynamic tasks are not known. In particular, the feature probability density functions are mostly unknown, and thus distribution free-classification must be performed. Consequently, adaptive neural networks are an attractive choice for classifying mental workload in complex real-world situations.

Techniques such as linear discriminant analysis (LDA) have been used for decades (Duda, Hart, and Stork, 2001; Bishop, 1995). However, as discussed previously, most real-world human cognitive and performance problems are not Gaussian in nature (Anderson, Devulapalli, and Stolz, 1995), and linear techniques may not provide adequate results. Other algorithms, such as support vector machines developed in the

1970's (Vapnik, 1999), have emerged as alternatives to the usual multilayer perceptron ANNs and discriminant analysis. With ANNs, the model classes are not restricted to linear input-output maps and the parameters are data-driven so as to match the model capacity to the data complexity. Support vector machines are an attractive alternative to the ANN since the data is linearly separable after a kernel transformation.

1.2.3 Adaptive Automation

Adaptive automation is the ability of the system to adapt to changes in operator cognitive demand and task performance and operator ability to respond to the situation (Freeman, Mikulka, Prinzel, and Scerbo, 1999; Parasuraman, Mouloua, and Molloy, 1996). Adaptive automation must be reliable to improve operator performance. Effective adaptive automation provides information that aids in decision making; it delivers the proper feedback at the appropriate time. Adaptive aiding aims to improve performance of the overall human-machine system. It must improve the system over existing static systems and over systems that are fully automated (Hancock and Verwey, 1997; Parasuraman, 1997). Adaptive automation, however, is not necessary if a fully automated system provides the same performance improvement without degradation of mission success.

Integrating key areas of research is necessary for improving operator performance with adaptive automation based on operator functional state. Operator functional state must be accurately measured and classified using robust pattern classification algorithms. In turn, the operator functional state must drive the adaptive automation. The automation must be appropriate for the task at hand and delivered at the appropriate time to improve operator performance and reduce operator cognitive workload.

1.3 Organization of Dissertation

The remainder of this dissertation is organized into sections. Section II provides a literature review of the contributions of the various disciplines required for developing an adaptive aiding system using operator functional state. Section 2.2 is an overview of the operational system used in this research. The mission and contingency operations of the UCAV are reviewed, illustrating the necessity of adaptively aiding the UCAV operator to improve performance. Section 2.3 is a brief introduction to operator state estimation, and Section 2.4 outlines psychophysiological assessment - a necessary component of operator state estimation. The applications using electroencephalography and their impact on this research are also explored in Section 2.4. The introduction and background for adaptive automation are discussed in Section 2.5. Sections 2.6 through 2.12 review the pattern classification algorithms used in this research, including multilayer perceptron artificial neural networks, support vector machines, and discriminant analysis classifiers. Techniques for determining saliency or importance of input features as well as methods for comparing pattern classification algorithms are also included in these sections.

Section III describes the experiments, methods, and measures used in this research while Section IV contains results and analysis of these experiments. These results clearly show the significant improvements in operator performance using operator functional state in union with adaptive aiding. Additionally, the results of the classifier comparison are explored; they indicate that multilayer perceptrons outperform the other candidate algorithms. Section V discusses the results of this research and conclusions about the utility of operator functional state as an input to an adaptive aiding system.

Finally, Section VI concludes this dissertation with an overview of significant contributions and some ideas for future research in the area.

II. Literature Review

2.1 Introduction

This chapter provides a review of the relevant methods and literature and also includes a brief overview of uninhabited combat air vehicles (UCAVs), their mission, and some areas which may stress the operator. Operator state estimation methods are reviewed with special emphasis on operator state estimation using psychophysiological measures. Artificial neural networks, particularly multilayer perceptrons using backpropagation training, are reviewed and feature saliency methods are discussed. Two sections discuss classifiers used or proposed by other investigators; these classifiers are based on discriminant analysis and support vector machines. Finally, methods of comparing these classifiers are considered.

2.2 Uninhabited Combat Air Vehicles

The Department of Defense has proposed a fleet of uninhabited air vehicles (UAVs) capable of strike missions in the most dangerous combat situations (Air Force Scientific Advisory Board, 1996). These prototypes can reduce cost in manufacturing and aircrew (Barry and Zimet, 2001) and plans exist to have the UCAV fielded by 2010. UAVs such as the Predator and Global Hawk allow commanders to obtain up-to-date information and images about the battlefield without risking pilots or ground forces. Even before the successful deployment of a Hellfire weapon from a Predator in early 2001, the idea for a specialized combat-capable UAV was explored (Air Force Scientific Advisory Board, 1996). This exploration culminated in the UCAV shown in Figure 1.

Figure 1. The Boeing X-45A is a UCAV being developed under a joint effort of the Defense Advanced Research Projects Agency (DARPA), the United States Air Force, and the Boeing Phantom Works.

The primary objective of the UCAV program is to develop a system to conduct suppression of enemy air defenses (SEAD) effectively and other strike missions (Borge, 2003). The UCAV operator must make decisions about targets based on weapons payload, remaining fuel, and target priorities while maintaining minimal radar cross section for four UCAVs. Controlling these parameters can be a very demanding task. In a statement was made about the planned taxi route (Garner, 2002), one of the first USAF UCAV operators stated that it was easy to become task saturated.

The primary concept of operations for the UCAV is the SEAD mission - a coordinated attack on known defenses, such as surface-to-air missile sites, that are near or enroute to other critical targets. These other critical targets would be removed using

manned assets such as strike aircraft. The UCAV routes and target assignments are preplanned with waypoints designated for capturing synthetic aperture radar (SAR) images of the target area and optimum weapon release points.

Other targets may 'pop up.' They can be avoided by mission replanning enroute or be targeted and eliminated by one of the four UCAVs. Decisions of this type depend on many variables such as fuel status, weapon status, and time pressures associated with completing the assigned mission.

Another mission envisioned for the UCAV is reactive suppression. This mission is much like attacking the 'pop up' targets described previously. These targets can be mobile missile launchers or unknown permanent locations. The UCAVs loiter near or over suspected enemy target locations and wait for the targets to appear on their sensors. The UCAVs capture a SAR image of the target location, assign weapons to the targets, and then attack the targets directly.

2.3 Operator State Estimation

Operator state has four major components (Gaillard and Kramer, 2000): psycho-physiological assessment (cognitive workload), operator performance assessment, situation awareness assessment, and momentary mission requirements as shown in Figure 2. The primary component focus in this research is 'closing the loop' of the human-machine system using cognitive workload alone. However, models for each component are necessary for accurate operator state assessment and, in turn, for intelligent adaptive aiding. For example, an operator may be unaware of an imminent threat (i.e., lacks situational awareness), but perform assigned tasks and have cognitive activity that

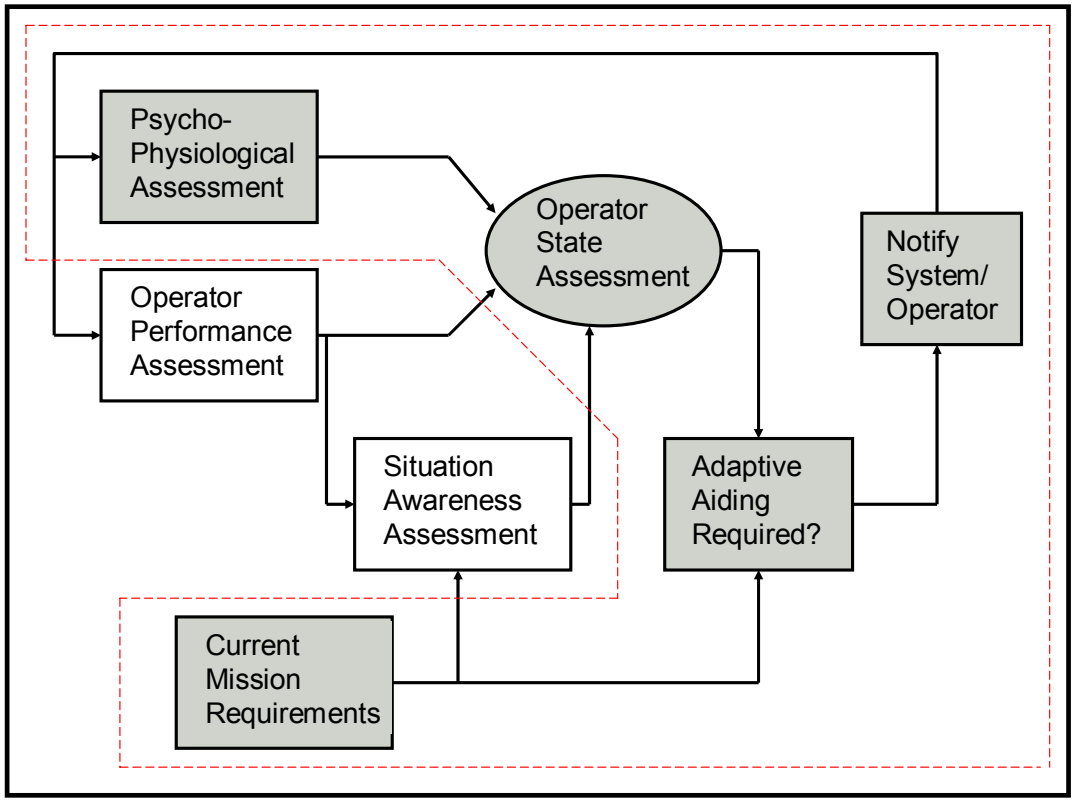

Figure 2. The Operator State Assessment Model with adaptive aiding consists of four major components for assessment of operator state. The components used in this research are highlighted and the system used is outlined by a dashed line.

indicates a normal or unstressed state. In this case, the components of operator functional state do not agree, and the operator should be notified of the impending threat.

2.4 Psychophysiological Assessment

The predominant and most obvious use of electroencephalography (EEG) is for clinical purposes. Less prominent uses include sleep research, brain computer interfaces, and research in classifying cognitive workload. Each of these areas of research is discussed in the following paragraphs, with emphasis on contributions to the assessment of operator cognitive load.

2.4.1 Clinical Research

Many studies have been conducted in the area of seizure detection using EEG. Most of these studies used wavelet and short-time Fourier transform techniques (Schiff, Aldroubi, Unser, and Sato, 1994) to identify the spikes evident during the onset of epileptic seizures, since classic spectral techniques do not contain the temporal information required to detect such spikes. Some of these studies used artificial neural network algorithms for online classification of epileptic spikes in background EEG (Galicki, Witte, Dörschel, Eiselt, and Griessbach, 1997; Szczuka and Wojdyłło, 2001; Liu, Zhang, and Yang, 2002).

Other clinical studies demonstrated the ability to classify abnormal and normal continuous EEG. These studies have included recognizing Alzheimer's disease (Pucci, Belardinelli, Cacchiò, Signorino, and Angeleri, 1999; Pritchard, Duke, Coburn, Moore, Tucker, Jann, and Hostetler, 1994; Petrosian, Prokhorov, Lajara-Nanson, and Schiffer, 2001) or Parkinson's disease (Robertson and Empson, 1999), and detecting pharmacological changes (Schaul, 1998; Gevins and Morgan, 1988), alcoholism (Winterer, Klöppel, Heinz, Ziller, Schmidt, and Herrmann, 1996), and psychosis (Szava, Valdes, Biscay, Galan, Bosch, Clark, and Jeminez, 1994; Kirsch, Bersthorn, Klein, Rindfleisch, and Olbrich, 2000; Hazarika, Chen, Tsoi, and Sergejew, 1997; John, Prichep, Fridman and Easton, 1988).

Classification of emotional state has also been demonstrated; for example, differences were detected between anger and happiness using psychophysiological measures (Waldstein, Kop, Schmidt, Haufler, Krantz, and Fox, 2000). Other researchers have suggested that personality can be detected in terms of convergent and divergent

thinking using EEG measures (Razoumnikova, 2000). The techniques and algorithms in clinical studies have crossed over into the other perspectives of EEG research as described in the following sections.

The methods used in clinical research referenced previously exhibit common techniques for classification and feature extraction. Similarities exist in the manner of signal processing of the EEG signal in clinical EEG research and other EEG research perspectives. The EEG is generally segmented into components based on frequency, and power measures are derived from average magnitudes within the frequency segments. These EEG segments have been labeled and frequency ranges for each EEG segment or band have been established. The bands are delta (~DC – 3 Hz), theta (4 – 7 Hz), alpha (8 – 12 Hz), beta (13 – 30 Hz), and gamma (31 – 42 Hz). Classification approaches are similar as well. Multivariate methods dominate the literature, but techniques using artificial neural networks are gaining acceptance.

2.4.2 Sleep Research

Early EEG recordings for sleep research (from the 1930s) were visually evaluated by clinicians since no automated methods of evaluating sleep signals were available (Uchida, Feinberg, March, Atsumi, and Maloney, 1999). With the advent of enabling technologies such as pattern recognition algorithms and appropriate computer hardware, clinicians are investigating automated techniques for determining sleep stages. For example, multivariate methods and power measures of EEG have been used to detect differences in Rapid Eye Movement (REM) sleep and other sleep stages (Uchida, Feinberg, March, Atsumi, and Maloney, 1999; Guevara, Lorenzo, Arce, Ramos, and Cori-Cabrera, 1995). Other researchers have used artificial neural networks and EEG to

classify sleep stages (Grözinger, Rösche, and Klöppel, 1995; Roberts and Tarassenko, 1992). Further studies examined awareness of auditory stimuli during drowsiness and sleep (Makeig and Jung, 1996) and used artificial neural networks to distinguish between alertness and drowsiness (Vuckovic, Radivojevic, Chen and Popovic, 2002). The techniques used are similar to those found in clinical research. Power measures are predominant, and the algorithms are consistent with those used in other applications.

2.4.3 Brain Computer Interface Research

A Brain Computer Interface (BCI) uses psychophysiological signals to control computer systems. For example, controlling a cursor on the screen using EEG measures is considered a BCI. Extensive work in the BCI area has suggested that this approach could be used as an alternate form of communication for severely handicapped persons (Keirn and Aunon, 1990; Keirn and Aunon, 1990). Algorithm development and classifier comparison have been investigated in imagined hand movements for control using multiple EEG channels (Pregenzer and Pfurtsceller, 1999; Ramoser, Müller-Gerking, and Pfurtscheller, 2000). Also, independent component analysis (ICA) and EEG have been investigated for control (Makeig, Enghoff, Jung, and Sejnowski, 2000).

Most of the literature in BCI research has been dedicated to measuring and detecting simulated hands and feet movements. Most research focuses on the use of EEG (Pregenzer and Pfurtsceller, 1999; Ramoser, Müller-Gerking, and Pfurtscheller, 2000; Keirn and Aunon, 1990, Müller-Gerking, Pfurtscheller, and Flyvbjerg, 2000; Peters, Pfurtscheller, Flyvbjerg, 1998; Polak and Kostov, 1997; Pfurtscheller, Neuper, Schlögl and Lugger, 1998; Peters, Pfurtschller, Flyvbjerg, 2001; Costa and Cabral, 2000; Mason and Birch, 2000), but some research has investigated the use of muscle activity

(Vaughan, Miner, McFarland, and Wolpaw, 1998) and combinations of EEG, muscle activity, and eye movement, including eye blinks (Russell and McMillan, 1999). Various classification algorithms have been investigated including multilayer perceptrons (Peters, Pfurtscheller, and Flyvbjerg, 1998; Peters, Pfurtscheller, and Flyvbjerg, 1998), committees of artificial neural networks (Peters, Pfurtscheller, and Flyvbjerg, 2001), tree-based neural networks (Ivanova, Pfurtscheller, and Andrew, 1995), time-delay neural networks (Haselsteiner and Pfurtscheller, 2000), Hidden Markov models (Obermaier, Guger, Neuper, and Pfurtscheller, 2001), min max modular neural networks (Lu, Shin, and Ichikawa, 2004), and linear discriminant analysis (Müller-Gerking, Pfurtscheller, and Flyvbjerg, 2000; Obermaier, Neuper, Guger, and Pfurtscheller, 2001; Millán, Mouriño, Franzé, Cincotti, Varsta, Heikkonen, and Babiloni, 2002). In BCI experiments, closed-loop real-time classification has been demonstrated using artificial neural networks and EEG measures (Guger, Schlögl, Neuper, Walterspacher, Strein, and Pfurtscheller, 2001; Guger, Ramoser, and Pfurtscheller, 2000).

BCI research represents the collection of requirements most similar to those necessary for cognitive load estimation and adaptive automation implementation. That is, reliable measures that are relatively simple to collect must be consistent across time and person. Also, real-time measurement and pattern classifiers must be developed to ensure accurate manipulation of the controlled systems.

2.4.4 Cognitive Load Estimation Research

Cognitive load is the mental activity associated with the performance of tasks. It has been assessed using central nervous system measures, such as continuous EEG as well as other psychophysiological measures, such as heart rate, eye blink, and eye

movement activity (Fournier, Wilson and Swain, 1999; Brookings, Wilson, and Swain, 1996; Wilson, Fullenkamp, and Davis, 1994; Wilson and Fisher, 1991, Wilson and Eggemeier, 1991). Cognitive or mental workload is considered high when the demands of the task challenge or exceed the capacity of the operator. Operator capacity can be affected by environmental factors such as heat, cold, noise, G-forces, etc., as well as individual factors such as fatigue, illness, and sleep loss (RTO Human Factors and Medicine Panel Task Group, 2004). High cognitive load can decrease operator performance and reduce operator awareness of new events or changes in events. As examples of physiological assessment of cognitive load research, the robustness of measures over time, the effects of learning, time pressure effects, and the effects of cognitive impairment are reviewed.

McEvoy, Smith, and Gevins (2000) examined robustness of measures over an hour and multiple day separation in data collection to evaluate the test-retest reliability of EEG signals as predictive measures. Task difficulty using EEG measures had high test-retest reliability in laboratory settings. The tasks examined were a working memory task and a psychomotor vigilance task. The data contaminated with muscle and eye movement artifacts was removed from analysis - usually impossible for a real-time classifier system since an answer is required regardless of contamination. In real-time systems, 'hand picking' data to determine cognitive state is not possible. Data collection for test and retest were separated by both one hour and approximately seven days. Pearson correlation coefficients showed significant reliabilities within session and between sessions with correlations above 0.9. Results showed that midline measures are better

than edge electrodes, since measurements from those electrode sites are less contaminated by muscle activity.

Learning effects cause differences in measurements of cognitive EEG activity. These effects are evident in new complex tasks, even if participants have previously experienced similar tasks (i.e., tracking targets with a mouse is not a new activity, but tracking targets with a mouse in a simulated ballistic missile attack is a novel task). The cognitive activity changes as the subject learns strategies for completing the imposed task. Changes in frontal theta (4-7 Hz) power and posterior alpha (8-12 Hz) power were found as participants developed strategies and learned the task (Smith, McEvoy, and Gevins, 1999). Other investigations found significant differences in eye blink rate and behavioral measures but could not find differences in EEG signals (Fournier, Wilson, and Swain, 1999).

In addition to learning effects, the effects of time pressure in a complex task results in differences in EEG activity (Slobounov, Fukada, Simon, Rearick, and Ray, 2000). As time pressure to complete a task increases, significant decreases in alpha (peak frequency of 10.5 Hz) power and increases in theta (4–7 Hz) and gamma (30-50 Hz) were found. This time pressure also caused performance breakdown, as indicated by an increased number of failed trials.

Cognitive impairment can be caused by many factors such as fatigue, sleep loss, hydration, circadian rhythms, and illness, and can cause changes in the 'normal' functioning of brain activity (Beaumont, Burov, Carter, Cheuvront, Sawka, Wilson, Van Orden, Hockey, Balkin and Gundel, 2004). For example, impairment due to intoxication or hangovers has been investigated using EEG (Gevins and Smith, 1999). Environmental

factors such as noise, vibration, sustained acceleration, and thermal stress also may affect cognitive activity (Fraser, Svensson, Grandt, Hockey, Balkin, Beaumont, Kamimori, Kautz, Belenky, Wesensten and Schlegel, 2004).

2.4.5 Pattern Classification Techniques for Cognitive Load Estimation

Many pattern classification techniques have been used to estimate operator functional state. Most prevalent are discriminant analysis (DA) techniques and artificial neural networks (ANN) as described in the following paragraphs. Support vector machines (SVM) have been used in brain computer interface research (Müller, Anderson, and Birch, 2003; Lal, Schröder, Hinterberger, Weston, Bogdan, Birbaumer, and Schölkopf, 2004; Garrett, Peterson, Anderson, and Thaut, 2003) but are not currently used in operator functional state estimation. Statistical process control with EEG measures has been used to classify pilot cognitive workload with limited success (Kudo, 2001).

Multivariate analysis techniques have been used in classification of cognitive workload research. Early research, enabled by the advent of faster and more readily available computers, used multivariate techniques for real-time processing of EEG data (Gevins and Morgan, 1986). Multivariate techniques have been used to classify levels of difficulty in a memory retention task (Wilson, Swain, and Ullsperger, 1999) and for determining levels of vigilance (Schober, Scellenberg, and Dimpfel, 1995). Stepwise discriminant analysis (SWDA) and ANNs were compared to classify pilot workload (Laine, Bauer, Lanning, Russell, and Wilson, 2002). Multivariate techniques were used to examine changes in EEG in simulated air traffic control (Brookings, Wilson, and Swain, 1996; Wilson, Swain, and Brookings, 1995), in simulated aviation tasks (Sterman, Mann,

Kaiser and Suyenobu, 1994), in actual flight tasks (Sterman and Mann, 1995; Wilson and Fisher, 1991; Wilson, Fullenkamp, and Davis, 1994), and in complex laboratory tasks (Smith, Gevins, Brown, Karnik, and Du, 2001; Wilson and Eggemeier, 1991).

Findings from these studies suggest that EEG measures can be used to determine multiple levels of cognitive load in complex tasks with results similar to those found in laboratory single-task experiments. Furthermore, the log power spectra EEG measures were sensitive to cognitive differences and reliable enough for consistent use, and allowing adequate time resolution for adaptive automation purposes. This finding is significant; laboratory tasks tend to be well structured and support consistent measurement of desired qualities. Complex tasks, however, tend to be less structured, and require operators to divide their mental capacity among several tasks.

Nontraditional measures have been evaluated for use in classifiers. Comparisons using coherence, cross phase, and cross power of multiple EEG channels and linear regression methods have been studied (Pleydell-Pearce, Whitecross, and Dickson, 2003; Valdés, Bosch, Graves, Hernandez, Riera, Pascual, and Biscay, 1992). Coherence and cross power of EEG have also been used with ANNs (Makeig, Jung, and Sejnowski, 1996). Interesting results of these studies included the use of coherence between EEG channels, which produced a dimensionless measure that maintained relational properties between channels. The use of independent component analysis for determining the source localization of individual EEG channels has been investigated with some success (Makeig, Bell, Jung, and Sejnowski, 1996). This method attempted to determine electrical signal sources within the brain from measures collected on the scalp.

ANNs have been used in a variety of EEG studies, often to automate continuous EEG analysis, thereby eliminating or reducing the need for visual inspection of the EEG recordings. This body of work can be categorized according to its purpose (Robert, Gaudy, Limoge, 2002): artifact processing, data compression, source localization, sleep research, clinical studies, cognitive workload studies, and brain computer interfaces.

Initial experiments using artificial neural networks to classify cognitive workload in complex tasks found that psychophysiological changes occurred before the onset of performance degradation in visiomotor memory tasks in fighter pilots (Gevins and Morgan, 1988). Differences were detected between alert and mentally fatigued pilots with 81 percent classification accuracy during long duration studies. Multilayer perceptrons with backpropagation training using eye blink and movement measurements were used to infer pilot workload by identifying flight segments (Siegel and Keller, 1992).

ANNs have also been used in the classification of cognitive workload in several studies including both simple single-task laboratory and complex multiple-task studies. The general use of artificial neural networks to classify differences in EEG has also been studied (Klöppel, 1994; Anderson, Devulapalli, and Stolz, 1995; Hazarika, Tsoi, and Sergejew, 1997; Gevins, Smith, Leong, McEvoy, Whitfield, Du and Rush (1998). Low, moderate, and high working memory load states were manipulated and each load pair in the classification process was compared. One group investigated single task workload classification using alpha band activity and autoregressive methods (Anderson, Devulapalli, and Stolz, 1995; Anderson, Stolz, and Shamsunder, 1998). Differences were detected between mental arithmetic and resting baseline using autoregressive models and

ANNs (Anderson, Stolz, and Shamsunder, 1995). Initial investigations using temporal and spatial information content were conducted using Elman recurrent ANNs (Greene, Bauer, Kabrisky, Rogers, and Wilson, 1997) with limited success. Cognitive workload estimation was investigated using EEG band activity and neural networks in simulated landing task (Russell, Monett and Wilson, 1996; Greene, Bauer, Kabrisky, Rogers, Russell and Wilson, 2000), in simulated air traffic control (Russell and Wilson, 1998; Wilson and Russell, 2003), in an air-to-ground Scud hunt mission (Russell, Reid and Vidulich, 2000), in complex laboratory tasks (Wilson and Russell, 2003), and for operators in a boiler plant simulation (Kurooka, Yamashita, and Nishitani, 2000). Classification accuracy varied for each of the studies but ranged from 70 to 98 percent. The results of these studies indicate that ANNs have been successfully used to accurately classify cognitive workload in a variety of environments.

2.5 Adaptive Automation

Most complex systems require the operator to adapt to changes in the environment or situation regardless of cognitive ability to accomplish required tasks in the changing environment. Adaptive automation is the ability of the system to adapt to changes in operator cognitive demand and task performance and operator ability to respond to the situation (Freeman, Mikulka, Prinzel, and Scerbo, 1999; Parasuraman, Mouloua, and Molloy, 1996). Adaptive automation must be reliable and must be provided when necessary to improve operator performance (Wilson, 2003; Parasuraman, 2003; Parasuraman, 1997). The key to automation is providing information that aids in decision making with the proper feedback at the appropriate time. Little research has been conducted to evaluate human capabilities in automation (Parasuraman, Sheridan,

and Wickens, 2000), but there is even less research that uses psychophysiological signals to control adaptive automation systems, especially in complex real world environments.

Adaptive aiding aims to improve performance of the overall human-machine system. It must improve the system over existing static systems as well as over fully automated systems (Hancock and Verwey, 1997; Parasuraman, 1997). If a fully automated system provides the same performance improvement without degradation of mission success, adaptive automation is unnecessary. Similarly, if upgrading existing systems, the adaptive automation must increase operator performance over the legacy static system (no automation). The aiding should provide an environment that fosters optimal human performance and prevent the operator from becoming overloaded, underloaded, or complacent. In both cases, operator performance may not be optimal. In some cases it may be disastrous. Consider the fighter pilot who is not aware of an enemy aircraft, the air traffic controller who manipulates so many aircraft that another aircraft entering assigned airspace is missed, or the truck driver on a long stretch of empty road who is not aware of a vehicle turning onto the road.

Another issue concerning adaptive automation is that the human operators themselves are adaptable and can respond to systems in unpredictable ways (Hancock and Verwey, 1997). Integration of system adaptive automation and natural human adaptation must be accomplished to eliminate the possibility of human-system instability. This integration may be accomplished by adding psychophysiological measures to the existing system (Prinzel, Freeman, Scerbo, Mikulka, and Pope, 1999; Byrne and Parasuraman, 1996). The operator cognitive state assessed by psychophysiological measures can be used as a control input to the system, adapting it only when the operator

is in a state of overload (Wilson, Lambert, and Russell, 2000). When the psycho-physiological measures indicate an increase in operator mental workload, the task or a group of subtasks can be automated, reducing mental demand on the operator.

Little research has been conducted using psychophysiological measures controlling closed-loop systems. However, single-task tracking experiments using EEG measures (Prinzel, Freeman, Scerbo, Mikulka, and Pope, 1999; Freeman, Mikulka, Prinzel, and Scerbo, 1999) have been conducted, and results showed significant improvements in operator performance with aiding. Aiding using human-computer communication tasks has also been investigated (Bubb-Lewis and Scerbo, 2002), and results indicated that aiding improved human-computer communication.

Wilson, Lambert, and Russell (2000) have conducted complex multiple-task laboratory experiments. The experiments consisted of multiple levels of workload using tracking, resource management, communications, and system monitoring tasks. The operators were aided when an increase in cognitive workload was detected using psychophysiological measures. The aiding consisted of full automation of communications and systems monitoring tasks. Adaptive aiding reduced tracking task error by 44% and resource management task error by 33%.

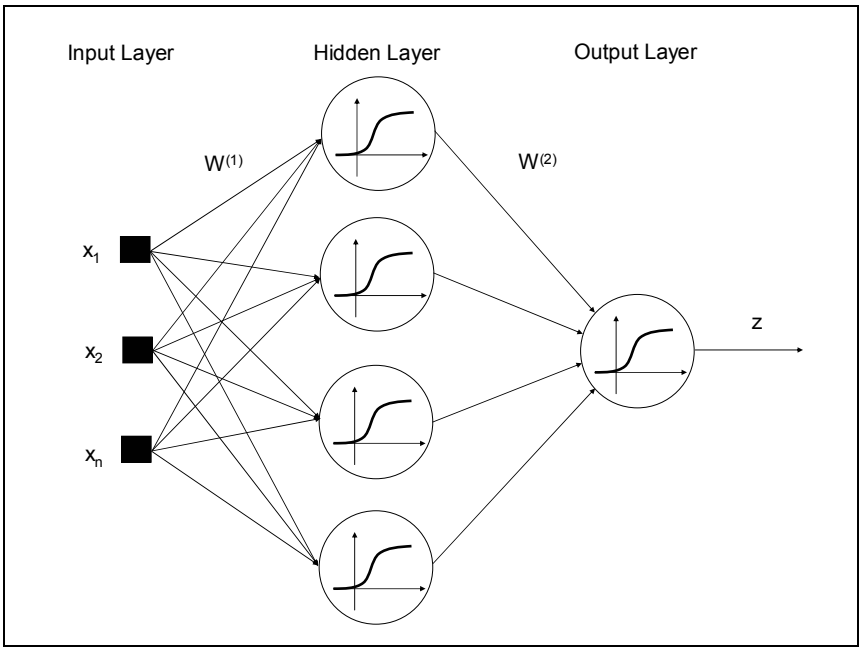

Figure 3. A fully connected multilayer perceptron ANN with inputs $x_1, x_2, \ldots, x_n$, output z, and layer weights $\mathbf{W} = \{\mathbf{W}^{(1)}, \mathbf{W}^{(2)}\}$.

2.6 Multilayer Perceptron Artificial Neural Networks

Feedforward multilayer perceptron artificial neural networks (ANN) with backpropagation training are among the most common ANNs for pattern classification applications (Widrow and Lehr, 1990; Lippmann, 1987). A mutilayer perceptron ANN classifier maps input vectors to output vectors in two phases. First, the network learns the input-output relationships from a set of training vectors that consist of input data (features) and the respective targets (assigned classes). Then, after training, the network acts as a classifier for new vectors.

Figure 3 shows the forward pass in addition to the fully connected feedfoward architecture of the multilayer perceptron, and Figure 4 shows a typical processing unit featuring summation and activation in a fully connected architecture. Each neuron in a

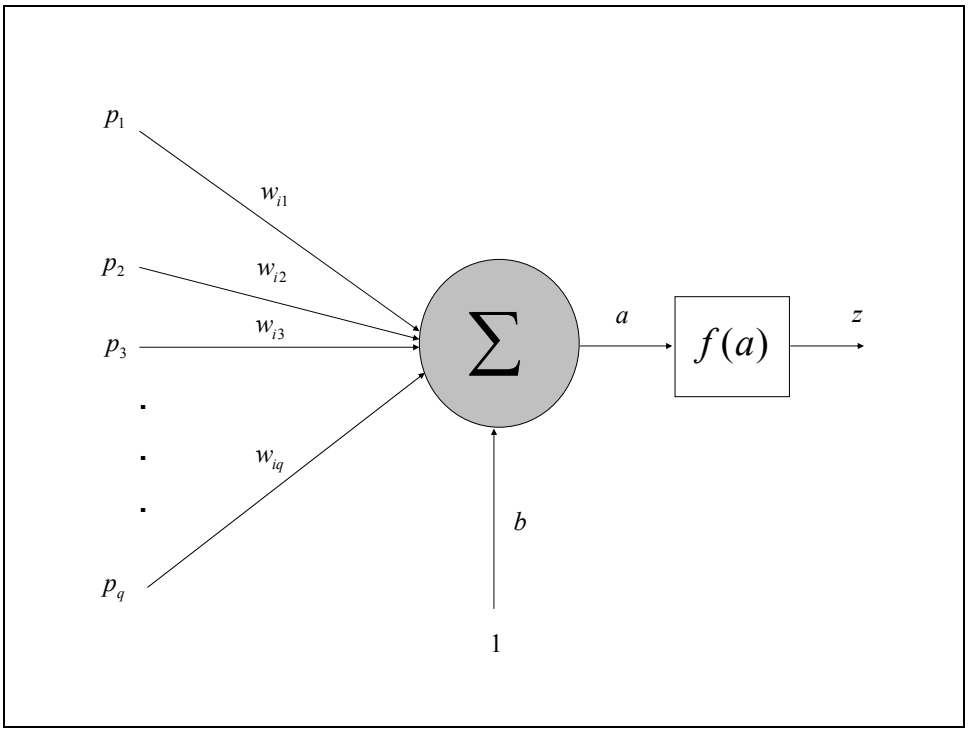

Figure 4. Individual neuron showing the weighted sum of inputs $a = \sum_{j=1}^{q} w_{ij} p_j + b$, followed by the logistic sigmoid activation function *f(a)* for neuron *i*, where *b* is a bias input.

layer is connected to every neuron in the preceding layer. The backpropagation algorithm initializes the network with a random set of weights for each fully connected layer, and then the network trains using given input-output pairs of training vectors. The 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, i.e. neurons not directly connected to any input or output. Each hidden unit then calculates the weighted sum of the input vector and its associated interconnection weights. Next, each hidden unit uses the weighted sum to calculate its activation that propagates to the output layer. Finally, each node in the output layer calculates its weighted sum and activation. The output of the

network is compared to the true output of the input-output pairs and their difference defines the output error.

In the second stage of backpropagation 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. Each hidden unit then calculates an error based on the error from each output unit. Next, the error from the hidden units is used to update the input weights. A single 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 are executed. The algorithm attempts to minimize the error or energy function

$$E = \sum_{k=1}^{m} \left| \vec{z}_k - \vec{t}_k \right|^2 , \tag{1}$$

where m is the size of the training set, $\vec{z}_k$ is the neural network output vector, and $\vec{t}_k$ is the true output (class) for each training input-output pair k.

The steps for implementing a feedfoward neural network with backpropagation training are as follows (Lippmann, 1987; Haykin, 1999; Widrow and Stearns, 1985; Widrow and Lehr, 1990):

(1) Initialize the weights w_l and biases b_l, where l is the current iteration.

(2) Present the input $\vec{p}_k$ and the target vector $\vec{t}_k$.

(3) Calculate the network output $\vec{z}_k$.

(4) Calculate the error E *(see Equation 1)*.

(5) Determine the new weights w_{l+1} where l+1 is the next iteration.

(6) Determine the new learning rate.

(7) Repeat steps 2 through 5 until desired error is achieved or when a predefined number of epochs are executed.

Each step is discussed individually in the remainder of this section.

Weights and biases are usually initialized with random numbers, often limited to the range –0.5 to 0.5, which is the nearly linear region of a sigmoidal activation function. This choice prevents the weights from starting in the extreme regions of the sigmoidal activation function, possibly increaseing training time. The maxima of the sigmoidal activation functions define the edges of the multidimensional error surface.

The data are usually normalized prior to presentation to the neural network, which prevents features with large magnitudes from dominating the learning and allows contributions from smaller and possibly more important features. The input data are normalized to zero mean and unit standard deviation using

$$p_n(i) = \frac{p(i) - \mu}{\sigma}, \tag{2}$$

where p_n is the normalized input vector, p is the input vector, μ and σ are the mean and standard deviation for each feature, and i represents the i^{th} training example.

Each input training vector is associated with a label defining the class to which that vector is assigned. The target vectors are assigned based on the labels defined *a priori*. Typically a vector is generated for each class as opposed to combining the target classes into a single output. Doing so would require applying a threshold to the output to determine the appropriate class. Thus, a target vector exists for each class that is assigned a high value, such as 0.9, if the data belongs to that class and a low value, such as 0.1, if it does not. For a two-class problem, the target vectors may be assigned $[0.9\ 0.1]^T$ for class 1 and $[0.1\ 0.9]^T$ for class 2.

The output of the network is determined by propagating the normalized input through each layer. As shown in Figure 4, the output of the individual node or neuron *j* is

$$z_i = f(a_i) \tag{3}$$

with

$$a_i = \sum_{j=1}^{q} \left(w_{ij} p_j + b_j \right), \tag{4}$$

where w_{ij} is the weight, p_j is the input, b_j is the bias and $f(a)$ is the activation function.

Activation functions can be linear or nonlinear. A common activation function is a sigmoidal nonlinearity (Haykin, 1999), usually a logistic sigmoid function with an output range $0 \le f(a) \le 1$ in the form

$$f(a) = \frac{1}{1 + e^{-a}}. \tag{5}$$

This activation function is chosen since it can produce the nonlinear hyperplanes required to classify data from most real-world applications.

The error is the difference between the output of the network and the expected target value as described by Equation (1). The weights are adjusted to minimize the error E_k through the backward path. Although the activation function is nonlinear, it is differentiable and $\frac{\partial E_k}{\partial w_{ij}}$ can be computed. The training algorithm is an extension of the Widrow-Hoff learning rule (Widrow and Lehr, 1990) - a gradient descent algorithm. This rule adjusts the weights using steepest descent, i.e.,

$$w_{ij}(n) = w_{ij}(n-1) - \eta \frac{\partial E}{\partial w_{ij}}, \tag{6}$$

where η is a learning rule constant that controls the speed of convergence at iteration *n*.

Adaptive learning and momentum are used to decrease the required training time. Typically, gradient descent methods use a fixed learning rate to control the rate of convergence (Widrow and Stearns, 1990). 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, incremental steps along the error surface are small and the algorithm is slow to converge to the desired error. Adapting the learning rate to optimize the learning progress maintains both stability and an acceptable rate of convergence. As the slope of the local error surface increases, the learning rate decreases to control stability.

Momentum helps to prevent the training algorithm from becoming trapped in a local minima (Haykin, 1999). Essentially the algorithm "jumps over" or ignores small perturbations in the error surface. Modification of the delta-learning rule to include momentum results in

$$w_{ij}(n) = \alpha w_{ij}(n-1) - \eta \frac{\partial E}{\partial w_{ij}}, \tag{7}$$

where α is the momentum.

The process repeats until a desired error is achieved. The desired error is problem specific and often determined by a cross-validation method that parses the data into three separate data sets: a training set, a validation set, and a test set. During training, the neural network adjusts the weights and biases based on the training set. After each adjustment the weights are tested on the validation set, and once the network reaches a minimum error, the test set is used to evaluate the final weights. The training and the validation error initially follow the same path until the neural network begins to learn the

idiosyncrasies of the training data set. The error for the training data continues to decrease after this point, but the validation error increases due to over-learning. The ideal stopping point for training is at the minimum validation error. Once trained, the weights are fixed and the network acts as a pattern classifier that examines input vectors it has never seen and predicts their class.

The number of nodes in the input layer, the hidden layer, and the output layer defines the architecture of the neural network. The number of input units and the number of output units are problem dependent. Typically, the number of neurons in the input layer is the number of features that form the full input space (Wilson and Russell, 2003). The output layer typically consists of the number of classes (Duda, Hart, and Stork, 2001; Wilson and Russell, 2003). 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 network with too few hidden units cannot learn the mapping to the required accuracy since the small hidden layer limits input space interaction. Too many hidden units allow the network to 'memorize' the training data so that it does not generalize well to new data. Typically, the size of the hidden layer is determined by training multiple multilayer perceptrons with different hidden layer sizes and then choosing the architecture with the best classification accuracy (Haykin, 1999).

2.7 Weight-based Partial Derivative Saliency Method

An important consideration in classification is selecting the input features. Some input features may be redundant because they are highly correlated or duplicated with

only scalar differences. Others may not provide useful information for discrimination. Decreasing the number of input features by removing redundant or meaningless inputs reduces the computation required for training. The "curse of dimensionality" abounds in pattern classification problems (Gershenfeld, 1999), including cognitive load state estimation. Psychophysiological signals collected in cognitive workload studies, such as EEG, electo-oculogram (EOG), and electrocardiogram (ECG), produce a gamut of derived features. As the number of input features increases, so do the number of training examples necessary to estimate the free parameters of the model.

Many approaches have been used to reduce the number of inputs by removing non-salient features. Among the most interesting are a weight-based partial derivative method (Ruck, Rogers, and Kabrisky, 1990) and a weight-based signal-to-noise ratio (SNR) method (Bauer, Alsing and Greene, 2000). Other approaches manipulate the inputs to reduce their number. Principal component analysis (PCA; Jolliffe, 1986; Flury, 1988; Dunteman, 1989) transforms correlated variables into uncorrelated variables. PCA determines the linear combinations for which the data have the maximum range of variability, thus reducing the number of variables. Each method presents different advantages and disadvantages as techniques for feature reduction. The PCA method will reduce the feature space for the classification algorithm but does not reduce the input space or the number of signals that must be collected. The partial derivative technique does not reduce the feature space by as much as the other two methods; however, it does provide a true input-output relationship for each feature. The signal-to-noise ratio method reduces both the input and feature spaces but requires a noise signal to inject into the classifier (Russell and Gustafson, 2001).

Another approach, the Ruck saliency measure (Ruck, Rogers and Kabrisky, 1990) determines which features provide information for classification by calculating the partial derivative for each network layer and ranking the features based on the saliency measure. This partial derivative method is possible because although the sigmoidal activation function or Equation (5) is nonlinear, it is differentiable, i.e.,

$$f'(a) = f(a)(1 - f(a)). \tag{8}$$

Feature saliency is based on the concept that a fully trained network contains all information for describing the relative importance of each input feature. Calculations are performed starting with the output layer whose partial derivative is

$$\gamma_{k3}^{(3)} = f'(a_{k3}^{(3)}) \tag{9}$$

$$= a_{k3}^{(3)}(1 - a_{k3}^{(3)}), \tag{10}$$

where k3 represents each output neuron and the superscript *(3)* denotes the third layer which is in this case the output layer. From Equation (4), *a* represents the weighted sum of the inputs to the activation function plus the bias or threshold. For the second or hidden layer

$$\gamma_{k2}^{(2)} = f'(a_{k2}^{(2)}) \sum_{k2} \gamma_{k3}^{(3)} w_{k2}^{(3)} \tag{11}$$

$$= a_{k2}^{(2)}(1 - a_{k2}^{(2)}) \sum_{k2} \gamma_{k3}^{(3)} w_{k2}^{(3)} . \tag{12}$$

where k2 represents the second layer neurons. For the input

$$\gamma_{k1}^{(1)} = f'(a_{k1}^{(1)}) \sum_{k1} \gamma_{k2}^{(2)} w_{k1}^{(2)} \tag{13}$$

$$= a_{k1}^{(1)}(1 - a_{k1}^{(1)}) \sum_{k1} \gamma_{k2}^{(2)} w_{k1}^{(2)} . \tag{14}$$

Finally, the partial derivative for the entire neural network is

$$\frac{\partial z_j}{\partial x_q} = \sum_{k1} \gamma_{k1}^{(1)} w_i^{(1)} \ . \tag{15}$$

Combining Equations (9) through (15) yields

$$\frac{\partial z_j}{\partial x_q} = \sum_{k1} \left[a_{k1}^{(1)} (1 - a_{k1}^{(1)}) \sum_{k1} \left[a_{k2}^{(2)} (1 - a_{k2}^{(2)}) \sum_{k2} \left[a_{k3}^{(3)} (1 - a_{k3}^{(3)}) \right] w_{k2}^{(3)} \right] w_{k1}^{(2)} \right] w_i^{(1)} \ . \tag{16}$$

Once the partial derivatives are calculated, the saliency is determined for each feature as

$$\Gamma_q = \sum_p \sum_j \left| \frac{\partial z_j}{\partial x_q} \right|, \tag{17}$$

where Γ_q is the saliency for the qth feature, j ranges over the outputs, and p ranges over the exemplar vectors in the training set.

The input features are rank ordered with features from largest to smallest saliency magnitude Γ_q. Features with the larger magnitudes contribute more toward separating the classes. Feature reduction can be accomplished by an iterative approach whereby a network is trained using all features, and the partial derivative saliency is calculated for each feature. The features are then rank ordered based on the computed saliency. The least salient feature is removed from the input matrix, the network is retrained using the reduced feature set and this procedure is repeated until all features have been removed from the training data set. The minimum data set is the smallest set that has acceptable classification accuracy. Figure 5 shows a typical response for this iterative process. The results are for 108 psychophysiological features from an air traffic control workload study which manipulated cognitive workload by increasing the number of aircraft monitored by the controller (Russell and Wilson, 1998).

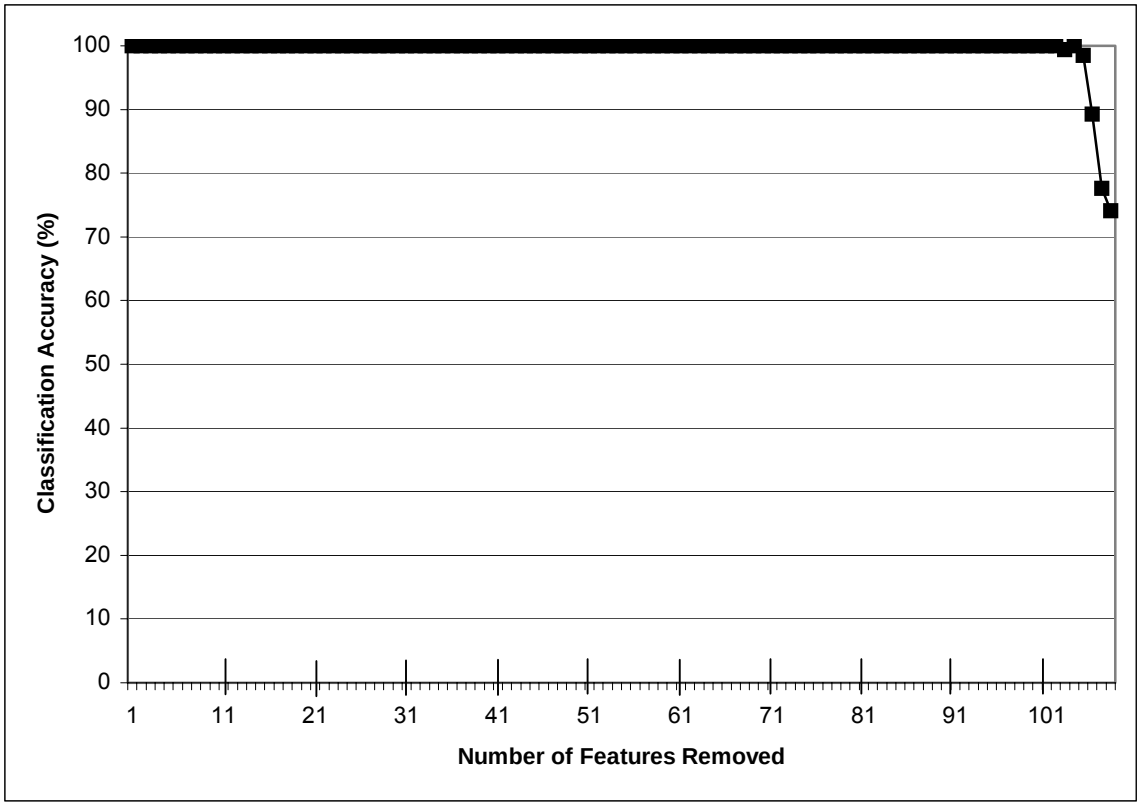

Figure 5. The classification accuracy remains nearly constant as non-salient features are removed, but accuracy decreases rapidly as salient features are removed.

2.8 Linear Discriminant Analysis

The classical technique of linear discriminant analysis was developed by Fisher in 1936 for two class problems and extended to multi-class problems by Rao in 1948 (Ripley, 1996). Fisher discriminant analysis performs dimensionality reduction while preserving as much of the class information as possible by maximizing the ratio of between-class variance to within-class variance (Duda, Hart, and Stork, 2001). Fisher discriminant analysis attempts to overcome the curse of dimensionality by reducing the number of dimensions before applying the classification algorithm (Bishop, 1995). The

dimensionality reduction to one dimension is accomplished by projecting the samples onto a line such that the values on the line are

$$y = w^T x\,, \tag{18}$$

where x is the sample vector and w is a vector of weight parameters. The values described by Equation (18) maximize the class separation and can be determined by adjusting the weight parameters w. An example of two projections of the same data, one optimal and one suboptimal, using different weight parameters is shown in Figures 6 and 7.

In Fisher discriminant analysis, the weight parameters are determined as follows. Let μ_r be the mean of data from class r,

$$\mu_r = \frac{1}{N_r} \sum_{x \in C_r} x\,, \tag{19}$$

where N_r is the number of samples in class r and C_r is the class to which the sample x_r is assigned. The mean of the projections for each class is

$$\tilde{\mu}_r = \frac{1}{N_r} \sum_{y \in Y_r} y\,. \tag{20}$$

Initially it may seem desirable to develop a distance measure that separates the means by substituting Equations (19) and (20) into Equation (18):

$$\left|\tilde{\mu}_1 - \tilde{\mu}_2\right| = \left|w^T\left(\mu_1 - \mu_2\right)\right|. \tag{21}$$

However, simply using the difference in the means may not produce the desired results.

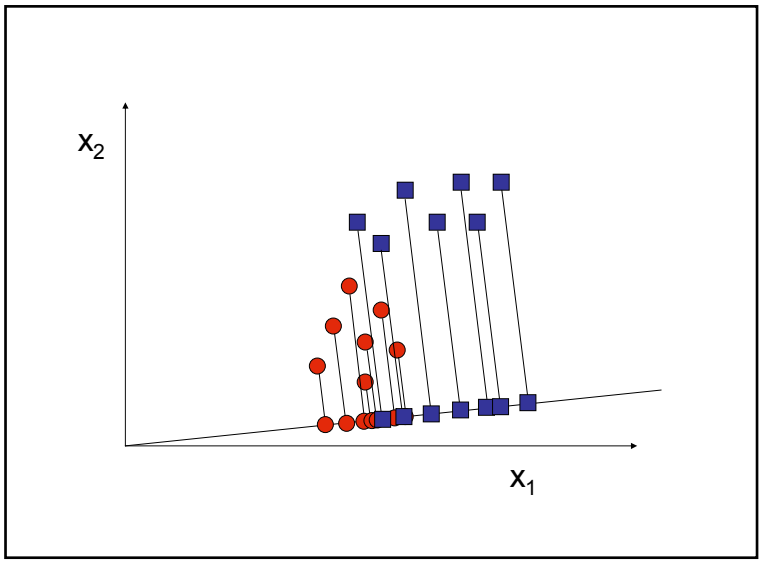

Figure 6. Projection of the samples onto a line using suboptimal weight parameters does not separate the two classes and is not optimal.

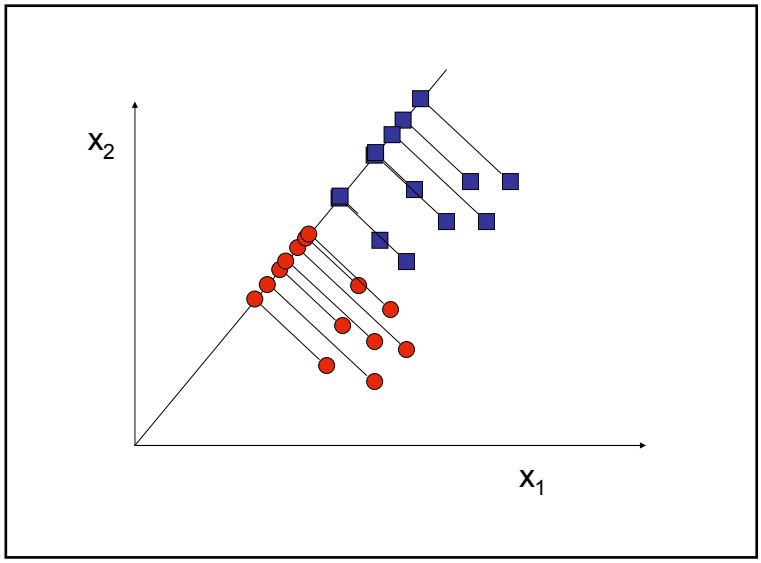

Figure 7. Projection of the samples onto a line that has optimal weight parameters yields good separation between the two classes.

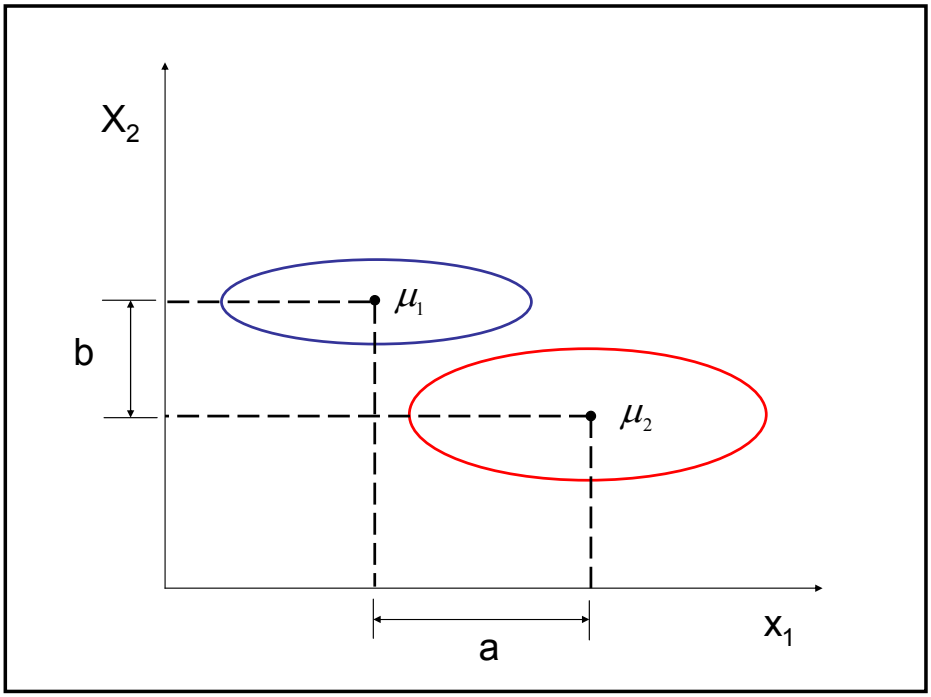

Figure 8. Two classes with means μ_1 and μ_2 have maximum separation *a* in the means for projection on the x_1 axis. However, greater class separation is achieved for projection on the x_2 axis, even though the separation *b* in the means is smaller.

For example, in Figure 8 the projection that yields the greatest separation in the means does not provide the best class separability (Bishop, 1995) because Equation (21) does not account for the variance of the classes.

Fisher's proposed solution maximizes a function that accounts for the separation in the means yet is normalized by a measure of the within-class scatter. To account for class variance, the class scatter for the projected samples is found (Bishop, 1995; Duda, Hart, and Stork, 2001), i.e.,

$$\widetilde{S}_r^2 = \sum_{y \in Y_r} (y - \mu_r)(y - \mu_r)^T \text{ , } r = 1, 2 \tag{22}$$

where the total within-class scatter of the projected samples is $\widetilde{S}_1^2 + \widetilde{S}_2^2$. The Fisher linear discriminant determines the w in $y = w^T x$ that maximize the Fisher criterion function

$$J(w) = \frac{\left|\widetilde{\mu}_1 - \widetilde{\mu}_2\right|^2}{\widetilde{S}_1^2 + \widetilde{S}_2^2}. \tag{23}$$

Maximizing this criterion determines a projection such that samples from the same class are projected close together and the projected class means are far apart.

The criterion function is in terms of the projected samples. As an explicit function of w, the criterion function must be in terms of the sample data x. The scatter or expected unnormalized covariance for each class is

$$S_r^2 = \sum_{x \in C_r} (x - \mu_r)(x - \mu_r)^T \text{ , } \tag{24}$$

where the within-class scatter is

$$S_W^2 = S_1^2 + S_2^2. \tag{25}$$

The scatter of the projection y can now be expressed as a function of the scatter matrix in terms of the feature space x. Substitution using Equations (18), (19), (21), (22) and (24) yields

$$\widetilde{S}_r^2 = \sum_{y \in Y_{ri}} (y - \mu_r)(y - \mu_r)^T \tag{26}$$

$$= \sum_{x \in C_r} \left(w^T x - w^T \mu_r\right)^2 \tag{27}$$

$$= \sum_{x \in C_r} w^T (x - \mu_r)(x - \mu_r)^T w \tag{28}$$

$$= w^T S_r^2 w \tag{29}$$

and

$$\tilde{S}_1^2 + \tilde{S}_2^2 = w^T S_W^2 w . \tag{30}$$

Similarly, the difference in the projected class means can be expressed in terms of the means in the original feature space and used to determine the between-class scatter:

$$\left(\tilde{\mu}_1 - \tilde{\mu}_2\right)^2 = \left(w^T \mu_1 - w^T \mu_2\right)^2 \tag{31}$$

$$= w^T \left(\mu_1 - \mu_2\right)\left(\mu_1 - \mu_2\right)^T w \tag{32}$$

$$= w^T S_B^2 w \tag{33}$$

where

$$S_B^2 = \left(\mu_1 - \mu_2\right)\left(\mu_1 - \mu_2\right)^T . \tag{34}$$

The Fisher criterion from Equation (23) can now be expressed in terms of the feature space using Equations (30) and (33):

$$J(w) = \frac{w^T S_B^2 w}{w^T S_W^2 w} . \tag{35}$$

Equation (35), the generalized Rayleigh quotient (Duda, Hart, and Stork, 2001), can be maximized by taking the derivative with respect to *w* and setting it equal to zero:

$$\frac{d}{dw}\left[J(w)\right] = \frac{d}{dw}\left[\frac{w^T S_B^2 w}{w^T S_W^2 w}\right] = 0 . \tag{36}$$

Using the chain rule yields

$$\left[w^T S_W w\right]\frac{d}{dw}\left[w^T S_B w\right] - \left[w^T S_B w\right]\frac{d}{dw}\left[w^T S_W w\right] = 0 \tag{37}$$

$$\left[w^T S_W w\right] 2 S_B w - \left[w^T S_B w\right] 2 S_W w = 0 \tag{38}$$

and dividing by $w^T S_W w$ yields

$$S_B w - \frac{w^T S_B w}{w^T S_W w} S_W w = 0 \tag{39}$$

$$S_B w - J(w) S_W w = 0 \tag{40}$$

$$S_W^{-1} S_B w - J(w) w = 0 . \tag{41}$$

Equation (41) is now a generalized eigenvalue problem in the form

$$S_W^{-1} S_B w = \lambda w , \tag{42}$$

where λ is the eigenvalue. Because only the direction of the data projection is important, solving for the eigenvalues is unnecessary (Bishop, 1995; Duda, Hart, and Stork, 2001) and the weights can be determined directly. Since $S_B w$ is always in the direction of $\mu_1 - \mu_2$, the solution is

$$w = S_W^{-1} \left(\mu_1 - \mu_2 \right). \tag{43}$$

Fisher discriminant analysis must also determine a threshold point along the one-dimensional subspace that separates the projected points (Ripley, 1996; Duda, Hart and Stork, 2001), i.e., the point along the projection where one class ends and the other begins. This threshold may be determined by modeling the projected data using normal probability densities and choosing the threshold w_0 as the point where the posterior probabilities of each class are equal (Bishop, 1995). The assignment of new data to each of the classes is then

$$\begin{array}{rl} w^T x + w_0 > 0 & \text{Class 1} \\ < 0 & \text{Class 2.} \end{array} \tag{44}$$

Generalizing Fisher discriminant analysis to multiple classes (linear discriminant analysis; LDA) is straightforward if the dimensionality of the input space is greater than or equal to the number of classes C (Bishop, 1995). LDA then produces C-1 projections

[$y_1, y_2, \ldots, y_{C-1}$] via C-1 projection vectors w_i, which can be arranged by columns into a projection matrix

$$W = [w_1 \mid w_2 \mid \ldots \mid w_{C-1}] \tag{45}$$

where

$$y_i = w_i^T x \Rightarrow y = W^T x . \tag{46}$$

A generalization of the within-class scatter (Equation (25)) is

$$S_W = \sum_{i=1}^{C} S_i . \tag{47}$$

A generalization of the between-class scatter is obtained using a total mean vector (Duda, Hart, and Stork, 2001)

$$\mu = \frac{1}{n} \sum_{\forall x} x = \frac{1}{n} \sum_{i=1}^{C} n_i \mu_i , \tag{48}$$

where n is the number of samples, and a total scatter matrix

$$S_T = \sum_{\forall x} (x - \mu)(x - \mu)^T , \tag{49}$$

or

$$S_T = S_W + S_B , \tag{50}$$

where

$$S_B = \sum_{i-1}^{C} n_i (\mu_i - \mu)(\mu_i - \mu)^T . \tag{51}$$

The criterion function from Equation (35) can now be written in terms of the multiclass S_W and S_B and the projection matrix W as

$$J(W) = \frac{\left| W^T S_B W \right|}{\left| W^T S_W W \right|} . \tag{52}$$

A scalar objective function is obtained using the indicated determinant. The projection matrix that maximizes the criterion function is found from a generalized eigenvalue problem by finding the roots of the characteristic polynomial

$$\left|S_B - \lambda_i S_W\right| = 0 \tag{53}$$

so that

$$\left(S_B - \lambda_i S_W\right) w_i = 0 \tag{54}$$

for each eigenvector. The largest eigenvalues indicate the directions of the greatest variance or spread of the data, i.e., the projections with the maximum class separability are the eigenvectors of $S_W^{-1} S_B$ with the largest eigenvalues (Bishop, 1995). Figure 9 shows relationships between the variables used in linear discriminant analysis.

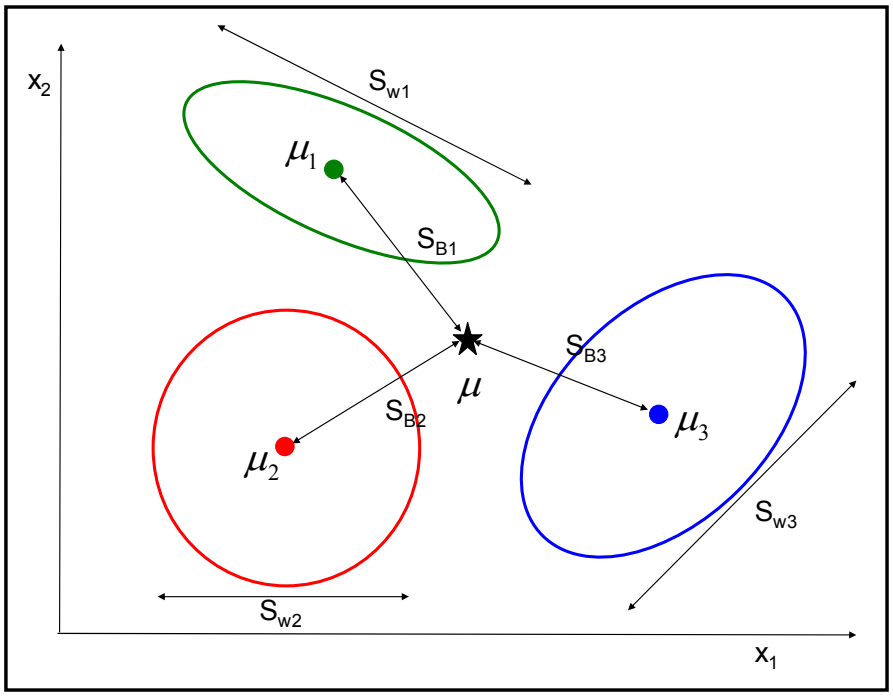

Figure 9. Linear discriminant analysis maximizes the ratio of between-class scatter S_B and within-class scatter S_W to define optimal linear hyperplanes for classification.

LDA can be derived as a maximum likelihood method for the case of normal class densities with equal covariance matrices (Fukanaga, 1990). LDA is optimal when the observations in each class have a multivariate normal density and each class has equal covariance matrices and equal prior probabilities. Two examples are explored here; both cases are three-class problems with class means $\mu_1 = [3 \quad 2]^T, \mu_2 = [7 \quad 4]^T, \mu_3 = [3 \quad 5]^T$. In the first case the covariance matrix for each class is $\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$. Figure 10 shows the probability density functions (assuming multivariate normal densities) for each class. Figure 11 is the probability density function plot rotated to project the densities to the x_1-x_2 plane. Since the covariance matrix is diagonal and the variance of

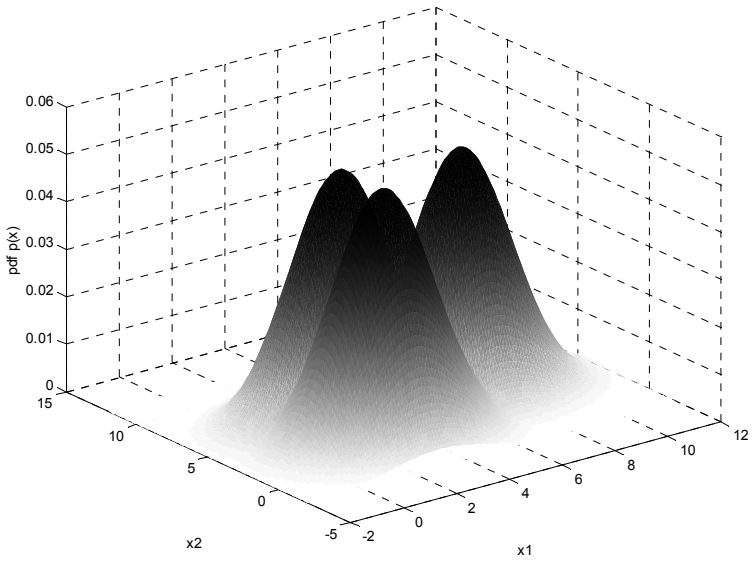

Figure 10. Probability density functions for three classes of data in two input variables, x_1 and x_2, with equal covariance.

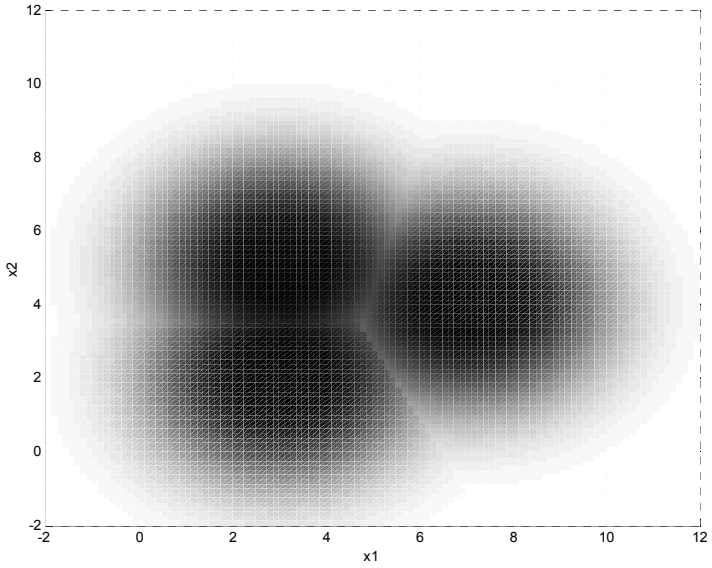

Figure 11. Projecting the probability densities onto the x_1-x_2 plane reveals optimal separating hyperplanes between the classes. Here each of the hyperplanes are lines for which the probability density functions of each class are equal.

each input is equal, the inputs are independent and the density contours are circular. The lighter shaded lines between the classes indicate where the probability of belonging to adjacent classes is equal.

Three sets of data are generated using the mean and variance parameters for the probability density functions described in Figure 10. Five hundred data points for each class are generated and presented to the linear discriminant analysis algorithm, and the results are shown in Figure 12. The separating hyperplanes are linear and map to the optimal lines displayed in Figure 11. For equal covariance matrices across classes, the linear discriminant analysis provides good separation between classes.

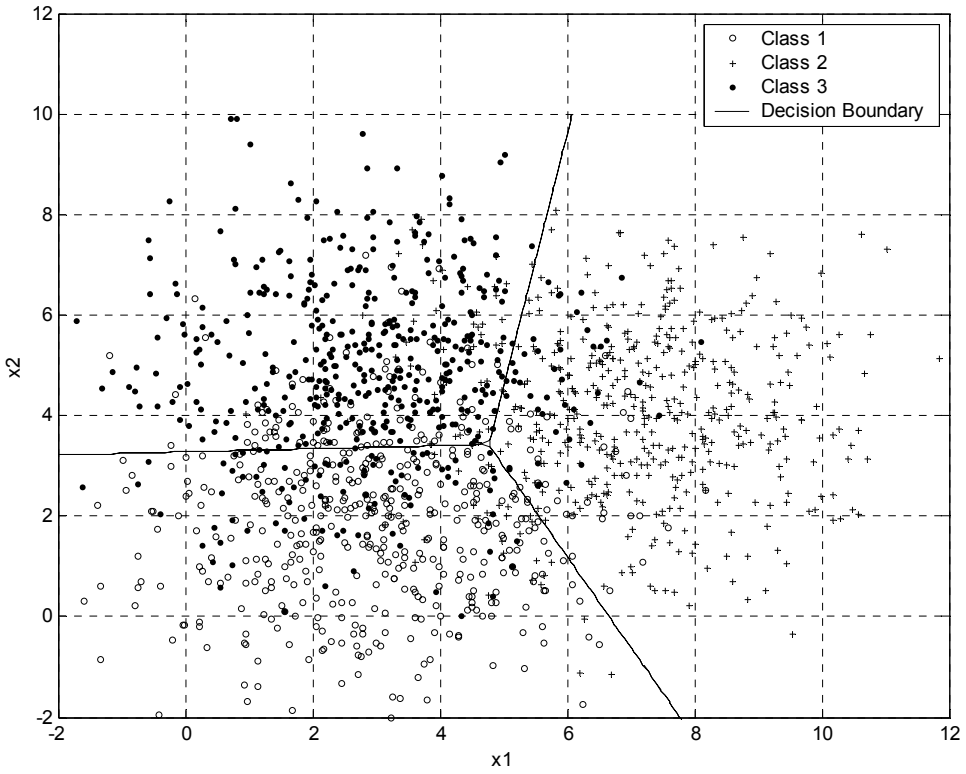

Figure 12. Randomly generated data in three classes are separated by linear decision boundaries. The class means are $\mu_1 = [3 \quad 2]^T, \mu_2 = [7 \quad 4]^T, \mu_3 = [3 \quad 5]^T$ and have equal across class covariance matrices, $\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$.

The second example uses unequal covariance matrices across the three classes. The means for the classes are as in the first example, but the covariance matrices across classes are not equal: $\Sigma_1 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$. Figure 13 shows the

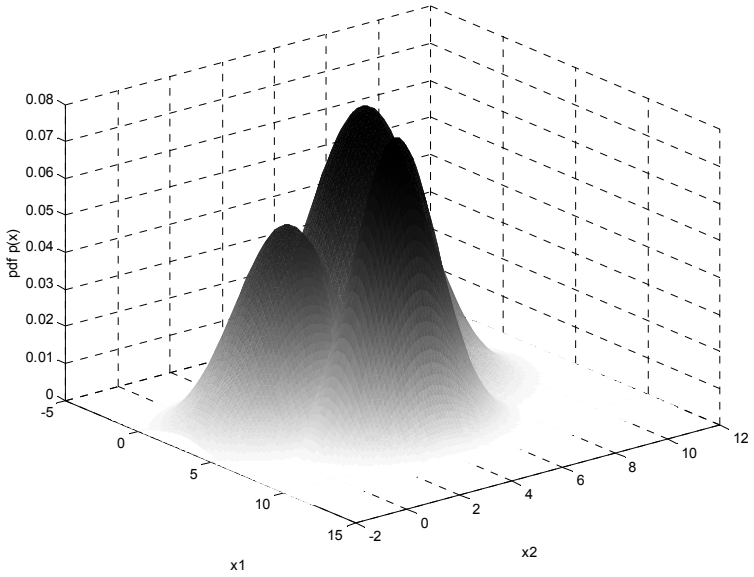

Figure 13. Probability density functions for three classes of data are displayed for two input variables, x_1 and x_2, with different covariance matrices.

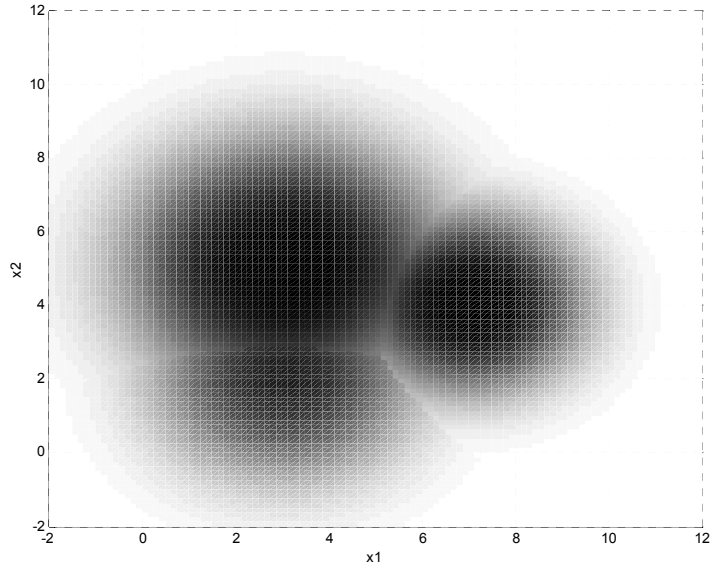

Figure 14. Projecting the probability densities onto the x_1-x_2 plane reveals the optimal separating boundaries between the classes. Each boundary indicates where the probability density functions of each class are equal.

probability density functions produced for this example, and Figure 14 shows the probability density function plot rotated to project the densities to the x_1-x_2 plane. The lighter shaded lines between the classes indicate where the probability of belonging to adjacent classes is equal.

Three sets of data are generated using the mean and variance parameters that define the probability density functions described in Figure 13. Five hundred data points for each class are generated and presented to the linear discriminant analysis algorithm,

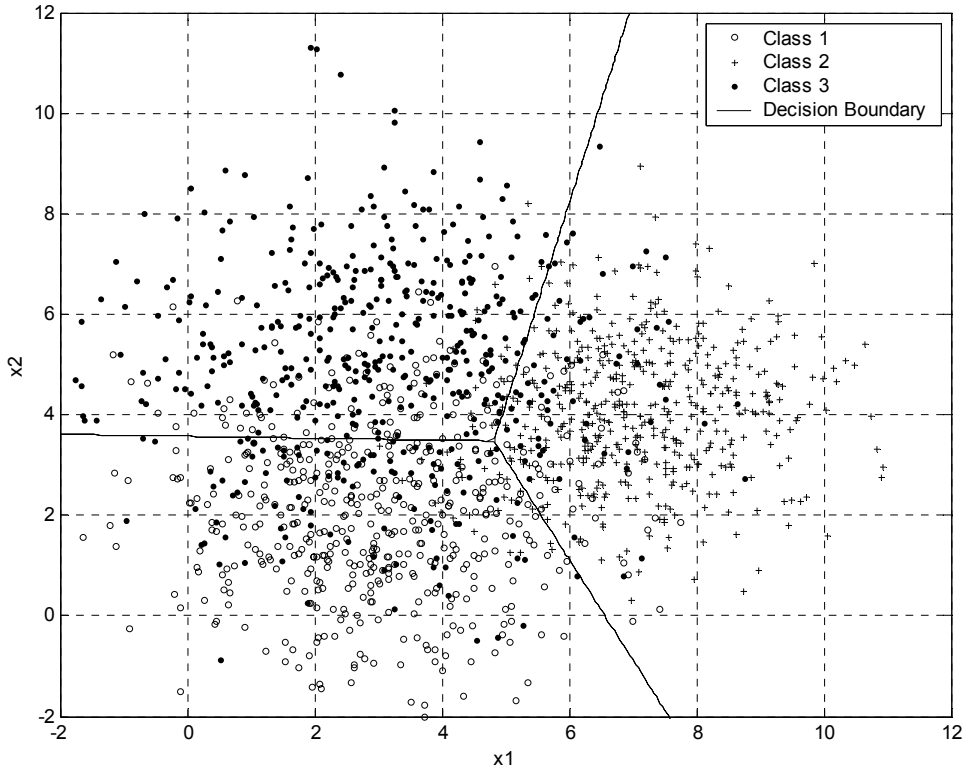

Figure 15. Randomly generated data consisting of three classes are separated by linear decision boundaries. The class means are $\mu_1 = [3 \quad 2]^T, \mu_2 = [7 \quad 4]^T, \mu_3 = [3 \quad 5]^T$ and the classes have unequal across class covariance matrices, $\Sigma_1 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, $\Sigma_3 = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$.

and the results are shown in Figure 15. The separating boundaries are linear and do not map to the optimal boundaries displayed in Figure 14. In the case of unequal covariance matrices across classes, linear discriminant analysis does not provide good separation between classes. The two examples illustrate that the important requirement for the LDA algorithm is equality of the covariance matrices.

The LDA algorithm does not perform well if the covariance matrices are not equal across classes and are only optimal for those cases (Fukanaga, 1990). Since the separating surfaces are not linear, unequal covariances will always require higher order input features to produce optimal separating hyperplanes. Quadratic discriminant analysis, as discussed in the next section, produces the required hyperplanes.

2.9 Quadratic Discriminant Analysis

Quadratic discriminant analysis (QDA) extends linear discriminant analysis (Fukunaga, 1990; Ripley, 1996) by including squared and cross products as well as linear functions of the predictor variables or features. The decision boundary in LDA is a linear function of the inputs; however, QDA produces a more flexible decision surface that is quadratic in the original measurement space but linear in the feature space (Hand, 1997). One approach that extends LDA to QDA transforms the inputs and does not assume an equal pooled covariance matrix, i.e., not $\sum = \sum_k$. A different approach used here transforms the inputs into a higher dimensional feature space. For two inputs, the transformation is $x_1, x_2 \rightarrow x_1, x_2, x_1 x_2, x_1^2, x_2^2$

The three sets of data generated for linear discriminant analysis using the mean and variance parameters that define the probability density functions of Figure 10 are presented to the QDA algorithm with results in Figure 16. The separating boundaries are

nearly linear and map to the optimal lines displayed in Figure 11. Increasing the number of samples for each class improves the model produced by quadratic discriminant analysis and ultimately leads to optimal lines.

The three sets of data generated for linear discriminant analysis using the mean and variance parameters that define the probability density functions described in Figure 13 are presented to the QDA algorithm with results in Figure 17. The separating

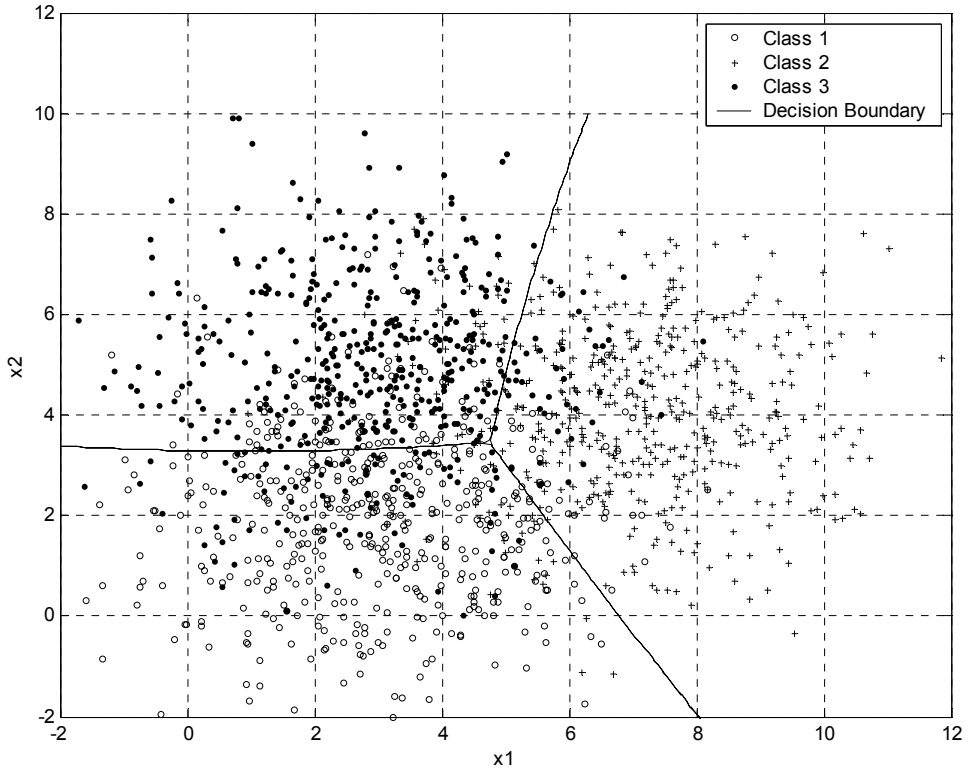

Figure 16. Randomly generated data consisting of three classes are separated by a linear decision boundary produced by quadratic discriminant analysis. The class means are $\mu_1 = [3 \quad 2]^T, \mu_2 = [7 \quad 4]^T, \mu_3 = [3 \quad 5]^T$ and have equal across class covariance matrices, $\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$.

boundaries are curvilinear or characterized by curved lines and map to the optimal curves displayed in Figure 14. As shown, QDA is superior to LDA for unequal covariance matrices across classes.

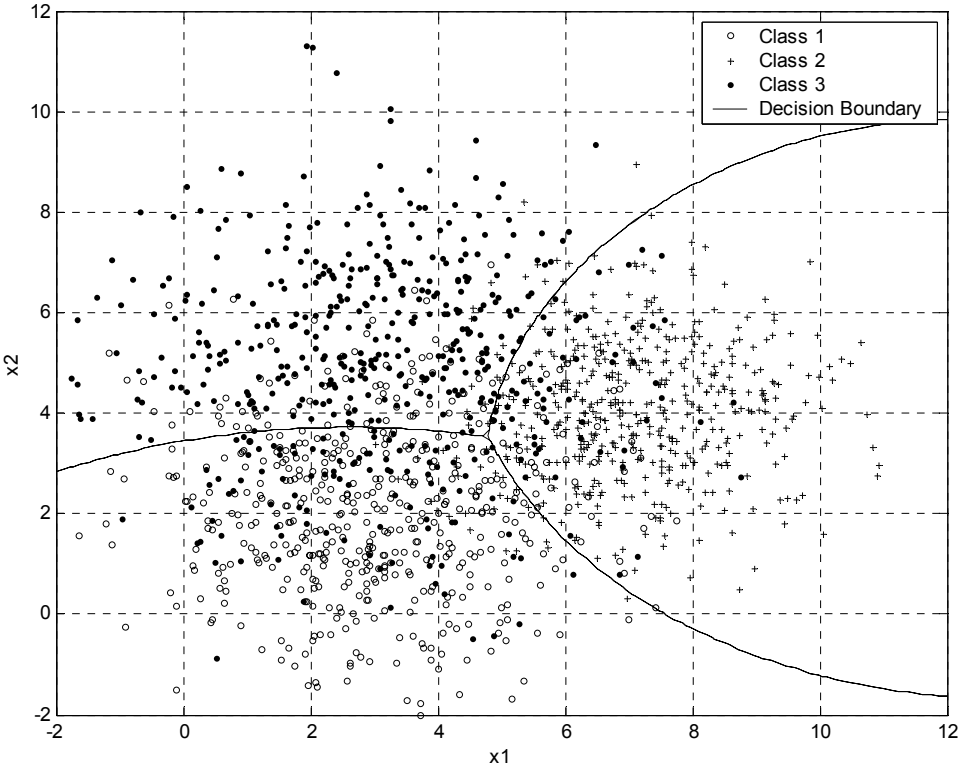

Figure 17. Randomly generated data consisting of three classes are separated by curvilinear decision boundaries produced by quadratic discriminant analysis. The class means are $\mu_1 = [3 \quad 2]^T, \mu_2 = [7 \quad 4]^T, \mu_3 = [3 \quad 5]^T$ and have unequal across class covariance matrices, $\Sigma_1 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$.

2.10 Logistic Discriminant Analysis

Logistic discriminant analysis or logistic regression analysis, a well known technique for classification, uses linear classification after a transformation (Ripley, 1996; Bishop, 1995). Unlike linear discriminant analysis, logistic discrimination does not assume class-wise Gaussian distributions. The only distributional assumption with this method is that the log likelihood ratio of the class distributions is linear in the observations. Further, this assumption is satisfied for a large range of exponential density families, e.g., Gaussian, beta, gamma, etc.

Logistic discriminant analysis uses estimates of the conditional posterior probabilities *Pr(C = k | X = x)* directly. C is the class and X is the input sample data since the class-wise distributions *f(C = k | X = x)* for class *k* given observation *x* and the prior probabilities (*Pr{C = k}*) are known and model the class posteriors in terms of K-1 log ratios (Ripley, 1996; Neter, Kutner, Nachtsheim, and Wasserman, 1996; Casella and Berger, 2002):

$$\log \frac{\Pr\{C = k \mid X = x\}}{\Pr\{C = K \mid X = x\}} = \beta_{k0} + \beta_k^T x, \quad k = 1,..., K-1, \tag{55}$$

where β is a weighting parameter on *x* and K is the number of classes. Thus the boundaries between classes are defined by

$$\log \frac{\Pr\{C = 1 \mid X = x\}}{\Pr\{C = K \mid X = x\}} = \beta_{10} + \beta_1^T x \tag{56}$$

$$\log \frac{\Pr\{C = 2 \mid X = x\}}{\Pr\{C = K \mid X = x\}} = \beta_{20} + \beta_2^T x \tag{57}$$

$$\vdots$$

$$\log \frac{\Pr\{C = K-1 \mid X = x\}}{\Pr\{C = K \mid X = x\}} = \beta_{(K-1)0} + \beta_{(K-1)}^T x\,. \tag{58}$$

An advantage of using such a model is that the posterior probabilities can be found as a simple closed form solution

$$\Pr\{C = k \mid X = x\} = \frac{\exp(\beta_{k0} + \beta_k^T x)}{1 + \sum_{l=0}^{K-1} \exp(\beta_{l0} + \beta_l^T x)}, \quad k = 1,..., K-1. \tag{59}$$

A well-known way to determine the free parameters β and fit the model is to use the maximum likelihood method (Fukinaga, 1990); it determines the probability density function as the one that makes the observed values *X* most likely. This criterion is obtained by determining the value of the parameter vector θ that maximizes the likelihood function L(θ) (Scharf, 1985, Shanmugan and Breipohl, 1988). The logistic discriminant model reasonably assumes that the observations *X* are independent and that the objective function for this model is the likelihood function

$$L(\beta) = \prod_{x \in (C=k)} \log \Pr(C = k \mid X = x; \beta), \quad k = 1,..., K-1. \tag{60}$$

The estimate of β that maximizes the likelihood function L(β) is the maximum likelihood estimator (Shanmugan and Breipohl, 1988). It is often easier to work with the log likelihood function

$$l(\beta) = \log L(\beta) = \sum_{x \in (C=k)} \Pr(C = k \mid X = x; \beta), \quad k = 1,..., K-1. \tag{61}$$

The same maximum likelihood estimate is obtained by maximizing either the likelihood or log likelihood functions since they are monotonically related.

Parameter estimation, however, is not as simple as the cases of linear discriminant analysis and quadratic discriminant analysis. Estimation must be accomplished using an iterative learning process such as a gradient-based method (Ripley, 1996; Duda, Hart, and Stork, 2001).

As an example of logistic discriminant analysis, consider two classes and binary classification (i.e., the output y is either 0 or 1). The boundary between the two classes is

$$\log\left\{\frac{P(X=x\mid C=1)}{P(X=x\mid C=2)}\right\}=\beta_0+\beta^T x\,. \tag{62}$$

Solving for the posterior probabilities yields

$$\Pr\{C=2\mid X=x\}=\frac{1}{1+\exp(\beta_0'+\beta^T x)} \tag{63}$$

and

$$\Pr\{C=1\mid X=x\}=\frac{\exp(\beta_0'+\beta^T x)}{1+\exp(\beta_0'+\beta^T x)}, \tag{64}$$

where

$$\beta_0'=\beta_0+\log\left\{\frac{\Pr(C=1)}{\Pr(C=2)}\right\}. \tag{65}$$

The likelihood and the log likelihood functions are

$$L(\beta)=\prod_{x\in(C=1)}\Pr(C=1\mid X=x)\prod_{x\in(C=2)}\Pr(C=2\mid X=x) \tag{66}$$

and

$$l(\beta)=\log(L(\beta))=\sum_{x\in C=1}\log(\Pr(C=1\mid X=x))+\sum_{x\in C=2}\log(\Pr(C=2\mid X=x)) \tag{67}$$

$$=\sum_{x\in C=1}\left(\beta_0'+\beta^T x\right)+\sum_{\forall x}\log\left\{1+\exp\left(\beta_0'+\beta^T x\right)\right\}. \tag{68}$$

Maximizing the log likelihood function requires an iterative learning process such as the Newton-Raphson algorithm, which uses partial derivatives with respect to the parameter vector β. The first and second derivatives are

$$\frac{\partial l(\beta)}{\partial\beta}=\sum_{\forall x}x\left(y-\exp\left(\beta_0'+\beta^T x\right)\right) \tag{69}$$

and

$$\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta^T} = -\sum_{\forall x} xx^T \exp\left(\beta_0' + \beta^T x\right)\left(1 + \exp\left(\beta_0' + \beta^T x\right)\right). \tag{70}$$

The estimates of β are updated using

$$\beta^{new} = \beta^{old} - \left(\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta^T}\right)^{-1} \frac{\partial l(\beta)}{\partial\beta} \tag{71}$$

until the difference between β^{new} and β^{old} is sufficiently small.

An alternate view considers logistic discriminant analysis as a nonlinear transformation of a linear combination of inputs (i.e., a transformation on the output of a linear summation) or

$$y = g\left(\beta_0 + \beta^T x\right) \tag{72}$$

where $g(\bullet)$ is the logistic transformation (Bishop, 1995). This view of logistic discriminant analysis is also the foundation of a single perceptron described in the artificial neural network and support vector machine sections, and is shown in Figure 18.

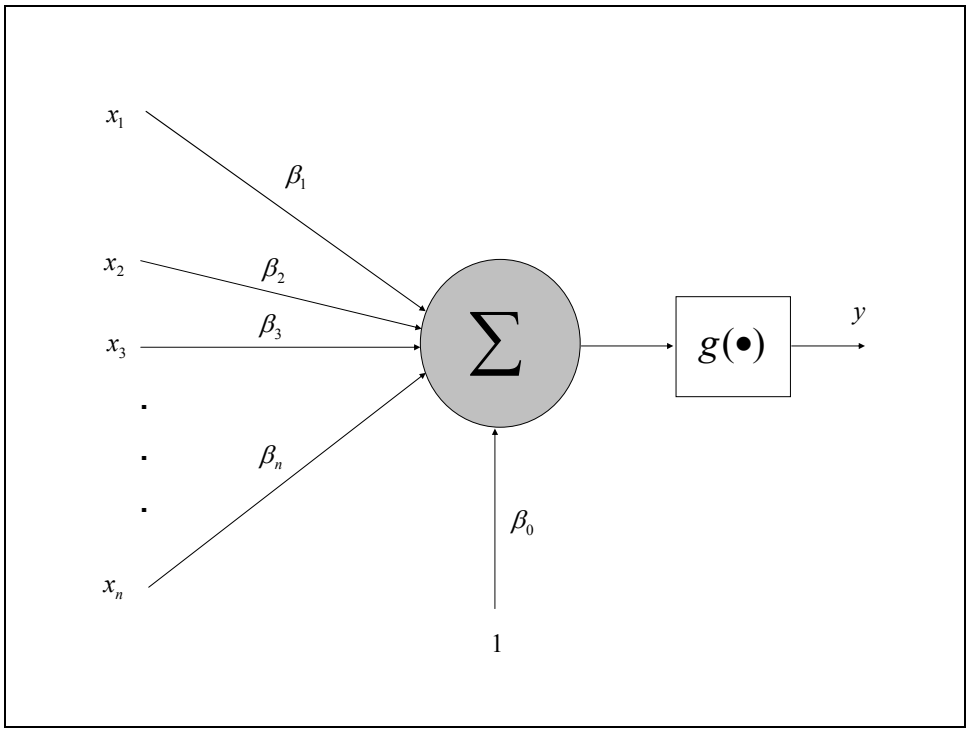

Figure 18. Logistic discriminant analysis may be considered a nonlinear transformation on a weighted summation of input variables similar to the perceptron.

2.11 Support Vector Machines

Kernel based learning algorithms, such as support vector machines, are basically comprised of two parts: a general learning machine and a problem specific kernel function (Vapnik, 1995; Burges, 1998). The support vector machine first transforms or maps the input data into a linear space using a kernel function and then applies a general learning machine to find the separating hyperplane. Support vector machines allow for model complexity as well as simplicity in model analysis. Multilayer perceptrons, radial basis function networks, and polynomial classifiers may be considered special cases of support vector machines (Müller, Mika, Rätsch, Tsuda, and Schölkopf, 2001). All have feedforward architectures as shown in Figure 19.

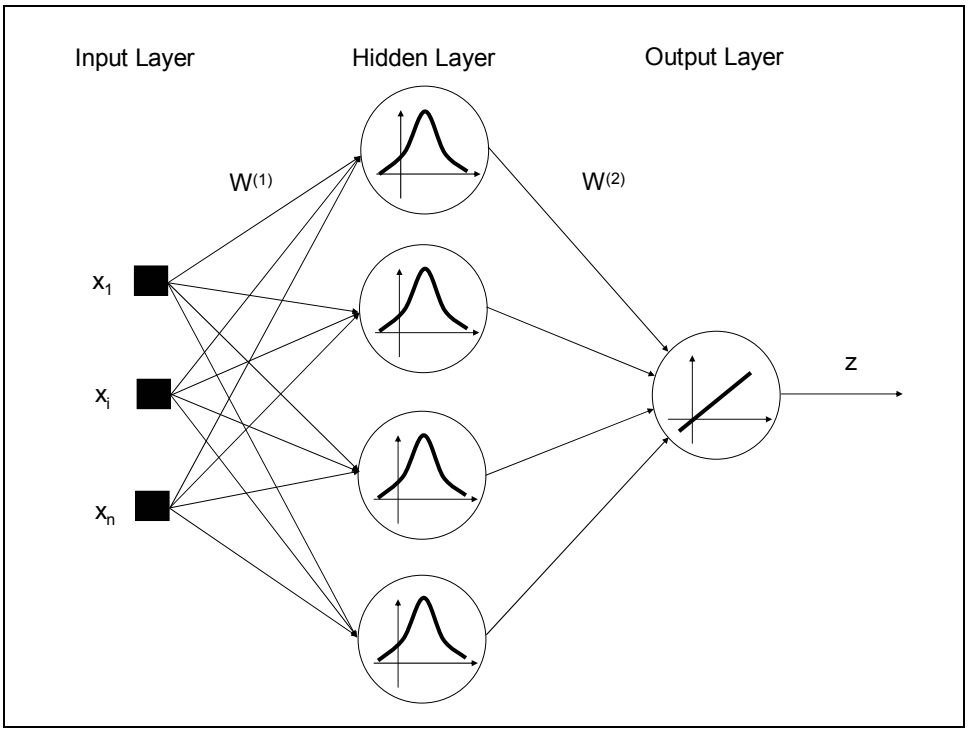

Figure 19. The support vector machine has the same feedforward architecture as most artificial neural networks. The important distinction is the learning algorithm.

Support vector machines map a nonlinear input space to a linear feature space using a kernel function and apply a linear algorithm to determine the hyperplane separating the classes. No computations are necessary in the high-dimensional input space. Kernel functions allow all computations to be performed in the linear feature space and permit quadratic optimization to produce an optimal separating hyperplane. Support vector machines provide good generalization by maximizing machine performance and minimizing model complexity simultaneously. These steps produce a support vector machine for classification:

1) Transform the input vectors into the feature space using an inner product kernel.
2) Determine the support vectors.

3) Compute the optimal separating hyperplane using quadratic optimization.

The perceptron, developed in the late 1950's, is one of the earliest artificial neural networks (Haykin, 1999, Duda, Hart and Stork, 2001, Bishop, 1995) and illustrates the support vector machine concept. This single-layer network has hard-limiting threshold activation functions that produce a 0 or 1 output providing linear separation of the input space as shown in Figure 20.

The hyperplane for the perceptron is defined by $f(x) = \langle w, x \rangle + b$ which is an inner product of the weight and input vectors. The inner product between vectors is

$$\langle \vec{x}, \vec{y} \rangle = \sum_i x_i y_i \ . \tag{73}$$

The activation function is the hard limiter or φ(x) = sign(f(x)). Points lying in the decision area in the direction of the weight vector are assigned a 1; those on the other side

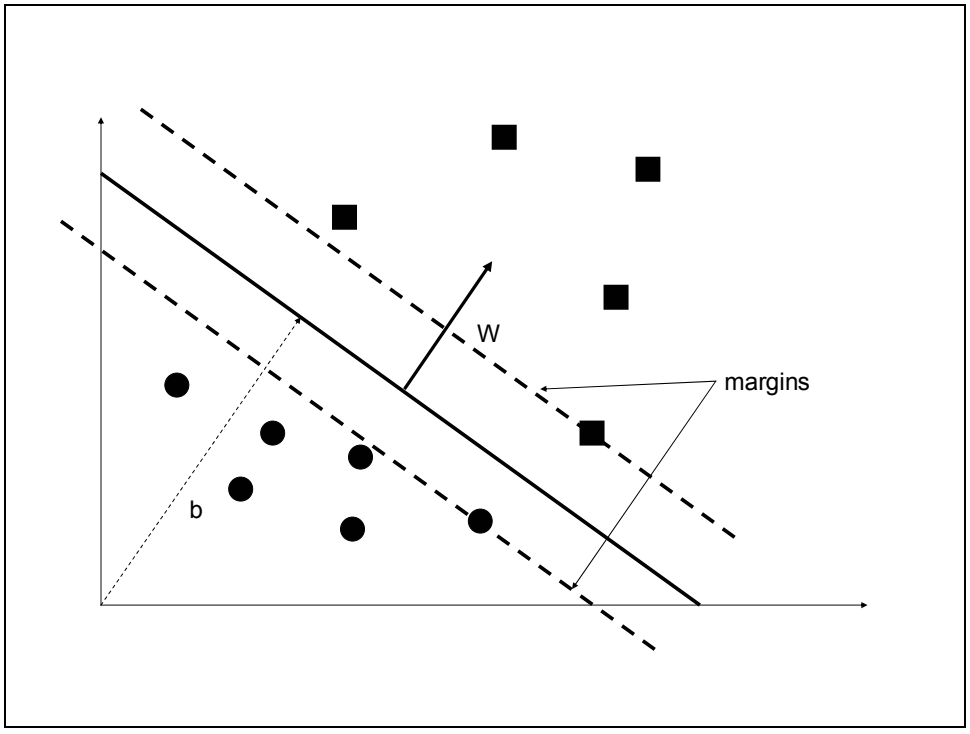

Figure 20. The perceptron defines a linear hyperplane which is the inner product of the weight and input space defined by $\langle W, x \rangle + b = 0$.

of the decision area are assigned a 0. The margins are the error bounds for particular data sets and are defined by the support vectors.

One advantage of using support vector machines over artificial neural networks is in the design of the architecture. Both have the same feedforward architecture, but training data determines the number of neurons in the hidden layer of the artificial neural network. This determination is significant. Selecting too few neurons results in poor classification (since the separating hyperplane is not well defined). Selecting too many neurons results in the risk of the classifier over learning the training data causing poor generalization.

2.11.1 Optimal Hyperplane Algorithm

Defining decision boundaries is a major difference between linear support vector machines and other linear methods for pattern classification. Linear discriminant analysis, for example, models the discriminant functions for each class as linear. Support vector machines model the boundaries between classes as linear.

Linear discriminant analysis and other classification methods define a hyperplane that separates the data (Figure 21). The hyperplane defined by these methods may not optimize the separation between the data and hence not optimize classification, particularly when the data are sparse. In linear discriminant analysis, the decision boundary is linear and defined by

$$y = W^T x . \tag{74}$$

Assuming the sample data set is sufficient (nonsingular), the pseudoinverse required to

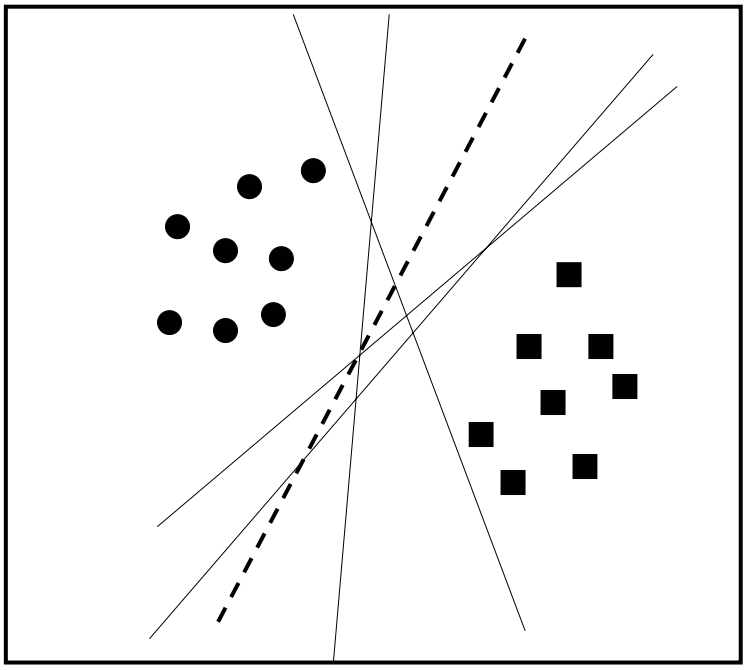

Figure 21. Many hyperplanes can be defined that completely separate the data, but only one optimally separates the data and evenly separates the data.

determine the parameters exists. A solution is provided if the data are Gaussian and parameter estimation is reduced to the minimization of the sum of errors squared. Another solution is to find an optimal hyperplane that maximizes the margin between the classes. The optimal hyperplane algorithm guarantees maximum separation with a maximum margin between the classes (Figure 22). The support vectors define the margins of the hyperplane, and the optimal hyperplane equally bisects the margins.

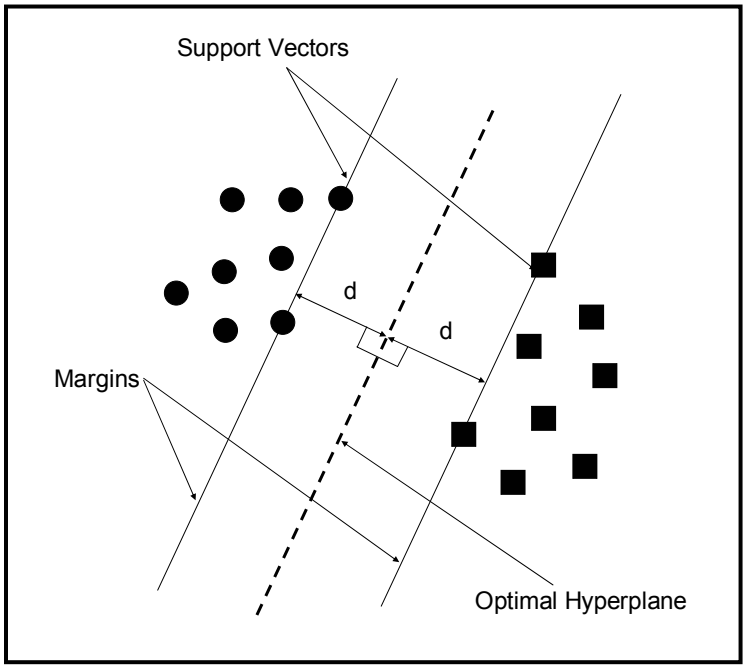

Figure 22. The optimal hyperplane maximizes the distance between all classes. The support vectors are those points on the margins.

2.11.2 High Dimensional Mapping and Inner Product Kernels

Kernel methods exploit the information contained in the inner products between data inputs as defined by Equation (73). Duality is the first condition required of inner product kernels for use in support vector machines. As previously shown in Section 2.11, the hyperplane for the perceptron is $f(x) = \langle w, x \rangle + b$, which is an inner product of the weight and input spaces. The solution is a linear combination of the training data,

$$w = \sum_i y_i x_i \,, \tag{75}$$

where y is the output vector and x is the input vector. The solution for the hyperplane has dual representation since it can be rewritten as

$$f(x) = \langle w, x \rangle + b = \sum_i y_i \langle x_i, x \rangle \,. \tag{76}$$

Note that in dual representation the data only appears inside the inner products.

Kernel methods map the nonlinear input space into a linear feature space. Data transformed nonlinearly into a high-dimension feature space is more likely to be linearly separable than in a lower dimension space (Cover, 1965). Support vector machines use kernel methods to map the lower dimension nonlinear input space into a linear high-dimension feature space (Figure 23). In accordance with Cover's theorem (Cover, 1965), the linear decision functions of the support vector machines should perform well in the high-dimension feature space.

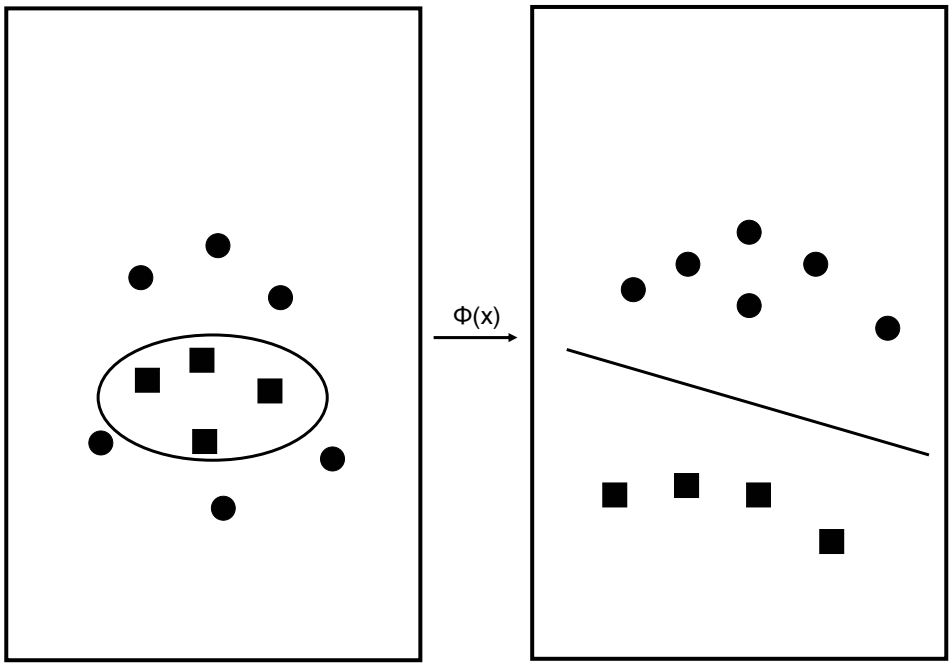

Figure 23. The kernel function maps a nonlinear input space (left) to a linear feature space (right).

A basic requirement for determining if a given kernel function is equivalent to an inner product in some space is based on Mercer's condition (Haykin, 1999; Vapnik, 1995). This condition must exist for a kernel function to map data to some other Hilbert space - a normed linear space with an inner product defined that is a generalization of Euclidean space (Scharf, 1985; Simmons, 1963). Mercer's condition states that there exists a mapping Φ and inner product expansion

$$K(x,y) = \sum_i \Phi(x)_i \Phi(y)_i \tag{77}$$

if, and only if, for any h(x) such that

$$\int h(x)^2 dx \text{ is finite,} \tag{78}$$

$$\int K(x,y)h(x)h(y)dxdy \geq 0\,. \tag{79}$$

Mercer's condition is sufficient to determine if a kernel is actually an inner product kernel in some space and can be used in a support vector machine. It says nothing on the techniques used to construct an inner product kernel.

Fortunately, several inner product kernels have been developed (Haykin, 1999; Vapnik, 1995). Two common ones for classification meet the criterion of Mercer's theorem: the polynomial learning machine, $\left(x^T x_i + 1\right)^p$ and the radial-basis function network, $\exp\left(-\frac{1}{2\sigma^2}\left\|x - x_i\right\|^2\right)$, where σ and p are specified parameters (Haykin, 1999). Additionally, the σ and p are *a priori,* problem-specific parameters that can be determined by experimentation using the data itself by varying the parameters and testing the classification results. For both support vector machine types, the number of support vectors extracted from the training data determines the dimensionality of the feature

space. The number of support vectors and their values determine the number of radial basis functions and their centers, respectively, in the case of the radial basis function support vector machines (Haykin, 1999).

2.12 Comparing Classifiers

Error rate is commonly used to compare classifiers. The error rate is

$$err = \frac{\sum_{\forall X} X_{missclassified}}{\sum_{\forall X} X}, \tag{80}$$

where $X_{misclassified}$ is an example mistakenly assigned to a wrong class and X is an example.

Confusion matrices are also used to evaluate classifier performance (Alsing, 2000). The confusion matrix and the truth table determine the within-class accuracy based on hits and misses. The truth table is simpler to compute and basically counts test samples from each class and the assigned class of those samples. Table 1 illustrates a sample truth table and shows that the classifier correctly assigned 450 class 1 test samples as class 1 but misclassified 50 class 1 test samples as class 2. Of the 500 samples from class 2, 450 were correctly assigned as class 2 and 50 samples were misclassified, with 25 samples assigned to class 1 and 25 samples assigned to class 3. All 500 samples from class 3 were correctly assigned to class 3.

The confusion matrix gives the class-conditional error rate (Ripley, 1996), i.e., it contains the posterior probabilities of a test sample assignment to each of the classes,

$$e_{ij} = \Pr\{\text{decision } j \mid \text{class } i\}. \tag{81}$$

Table 1. A truth table compares test classification counts with the truth. Rows indicate truth and the columns indicate the test result.

	Class 1	Class 2	Class 3
Class 1	450	50	0
Class 2	25	450	25
Class 3	0	0	500

Table 2. A confusion matrix shows the probability that new data from class 1 is classified as class j = 1, 2, 3.

	Class 1	Class 2	Class 3
Class 1	0.90	0.10	0.00
Class 2	0.05	0.90	0.05
Class 3	0.00	0.00	1.00

The confusion matrix in Table 2 shows the class-conditional probabilities for the example in Table 1. Class 1 was correctly predicted in 90% of the instances of the test data; class 1 was misclassified 10% of the time as class 2 but never misclassified as class 3. Likewise, class 2 was correctly predicted in 90% of the instances of the test data; however, 5% of the test samples were misclassified as class 1 and another 5% were misclassified as class 3. All test samples from class 3 were correctly classified and the classification prediction for class 3 was 100%.

Besides directly comparing classification accuracy, classifiers can be compared using error rates. Each sample to be tested is a discrete event with two possible outcomes: correct or incorrect. These independent, identical trials are Bernoulli trials with two possible outcomes (Casella and Berger, 2002). A series of these random Bernoulli trials has a binomial distribution. By comparing the number of successful trials, comparisons of competing classification algorithms can be made. A multinomial selection procedure uses

these comparisons to determine the best classification algorithm for a given test set (Alsing, 2000; Alsing, Bauer, and Miller, 2002). The multinomial selection procedure is as follows (Alsing, Bauer, and Miller, 2002):

1) Compare class posterior probabilities for each classifier.

2) Find the largest class posterior probability for each data point.

3) Determine which classifier has the largest posterior probability.

4) Compute the number or wins for each classifier.

5) Rank the wins.

6) Declare the classifier with the most wins to be the best classifier.

Another method of comparing classification algorithms is McNemar's test (Ripley, 1994). This method is similar to the multinomial selection procedure but compares classifiers pairwise. It uses the errors of each classifier, which also have a binomial distribution. McNemer's test is

$$M = \frac{|n_A - n_B| - 1}{\sqrt{n_A + n_B}}, \tag{82}$$

where n_A is the number of errors made by classifier A but not classifier B and n_B is the number of errors made by classifier B but not A.

The measure M can be compared to a chi-squared distribution with one degree of freedom as a test for the improvement in correct classification in classifier A versus classifier B (Schealler and McClave, 1986). The chi-squared probability of observing a value of n_A or less, given the null hypothesis of a binomial distribution, $B(n_A + n_B, 1/2)$, serves as a test for the improvement of the estimation in classifier A over classifier B.

The multinomial selection procedure and McNemar's test require that the classifiers use the same test data, but these tests are not intended as single measures for

determining classifier performance. For example, classifier A may show significant improvement over classifier B using the McNemar's test even though their classification accuracies differ only slightly. Algorithm complexity, ease of use, selectivity (classifier accuracy), and specificity (class posterior probabilities) must be considered when determining the best classifier to use in applications.

Each method has advantages and disadvantages. The most apparent disadvantage to each is that no single method completely describes the results of the classifier comparison. For example, the error rate does not provide information on the misclassifications; it only provides overall classification accuracy. The addition of confusion matrices provides model specificity in the class-conditional error. McNemar's test and the multinomial selection procedure provide tests for improvement in classification between classifiers. The multinomial selection procedure can determine the best classifier from many (two or more) while McNemar's test can only perform pairwise comparisons. Both tests provide no information on classifier specificity or selectivity and only determine which classifier provides the best results. For those reasons, combining results from multiple classifier comparison methods provides a more informative picture of the strength of classifier algorithms.

2.13 Section Summary

Section II introduced the foundational literature for this research. As such, classifier algorithms were explored. Background and supporting information on the measurement of psychophysiological data and its applications were reviewed, with emphasis placed on the methods used in this study. Adaptive automation for integration into human-machine systems was also explored. Finally, the military platform used in

this research was introduced. The next section describes the use of this background information in a complex military platform and describes the experimental methods and motivation for the methods.

III. Methodology

3.1 Methodology Overview

This section describes the experimental methods used in this research. The tasks performed by a UCAV operator and the psychophysiological measures gleaned from the literature review are discussed in considerable detail. Methods of data collection signal processing, and integration are outlined; new measures are presented, and methods for integrating these measures are described. The operator performance and subjective measures used in the experiments are also defined in this section.

To meet dissertation objectives, this research is based on two experiments. The first is a single-task experiment for developing multiple cognitive models derived from information processing demands and task type. The data from this experiment is also used to compare the classifier algorithms considered in this research. The second is a dual-task experiment for determining the mission effectiveness of adaptive aiding using operator functional state in an operationally relevant environment.

3.2 UCAV Research Platform

The UCAV simulator discussed in this research was developed by AFRL/HECI, System Control Interfaces Branch, to explore interface design and was modified by AFRL/HECP, Collaborative Interfaces Branch, to investigate real-time adaptive aiding techniques. It simulates the forward area of operations, i.e., the ingress and weapons delivery portion of the UCAV mission. Tasks include synthetic aperture radar (SAR) downloading and processing, setting designated mean points of impact (DMPIs), and authorizing, arming, and clearing weapons for release.

A single operator monitors four UCAVs during the simulation of a Suppression of Enemy Air Defense (SEAD) mission. The operator monitors the ingress of the four vehicles until they reach the SAR capture waypoint. Once the SAR is captured, the operator downloads the SAR image from the UCAV to the operator station, visually processes the SAR images, and selects DMPIs. After selecting the targets, the operator updates the shoot list, arms the weapons, and authorizes the release of weapons. The operator completes this process for all four UCAVs.

Figures 24 and 25 show the interface for the UCAV operator workstation. Figure 24 is during vehicle ingress to the targets, and Figure 25 shows the target selection process. The operator conducts all these tasks (selecting weapons, placing the DMPIs on the target, and authorizing the release of weapons) on the right side of the screen, and hereafter those tasks are collectively referred to as the operator vehicle interface (OVI) task.

Processing SAR images is a difficult task. The operator must locate targets regardless of target orientation and background clutter such as trees. Some targets may even be occluded by the background clutter. This study used three target types embedded in forest: Type A (communication and command and control trailers), Type B (SA-10 surface-to-air missiles), and Type C (SA-12 surface-to-air missiles). Type A and Types B and C targets were considered easy and hard to locate, respectively. Figure 26 shows examples of all three target types.

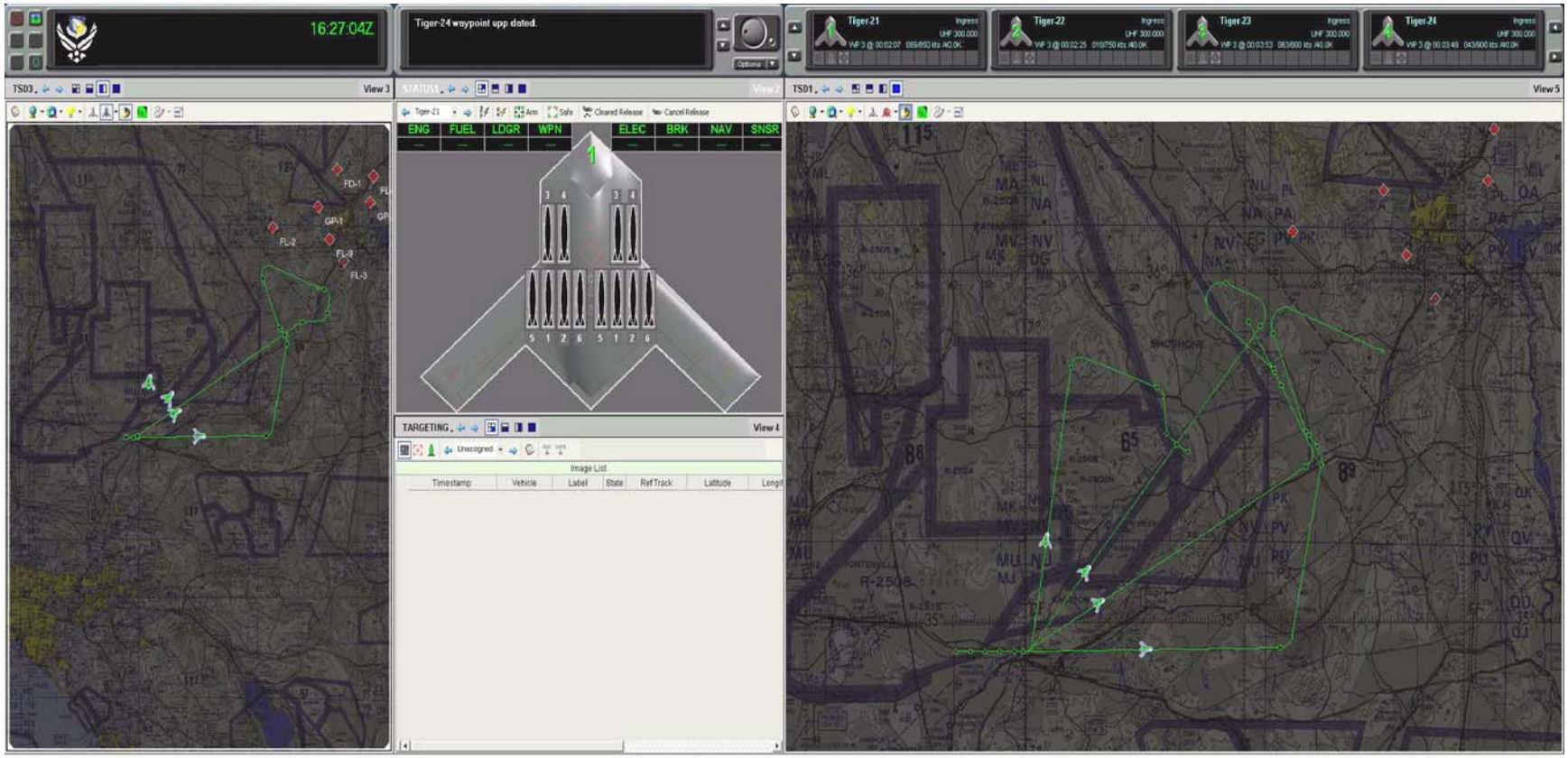

Figure 24. Sample UCAV display showing the ingress of four UCAVs .

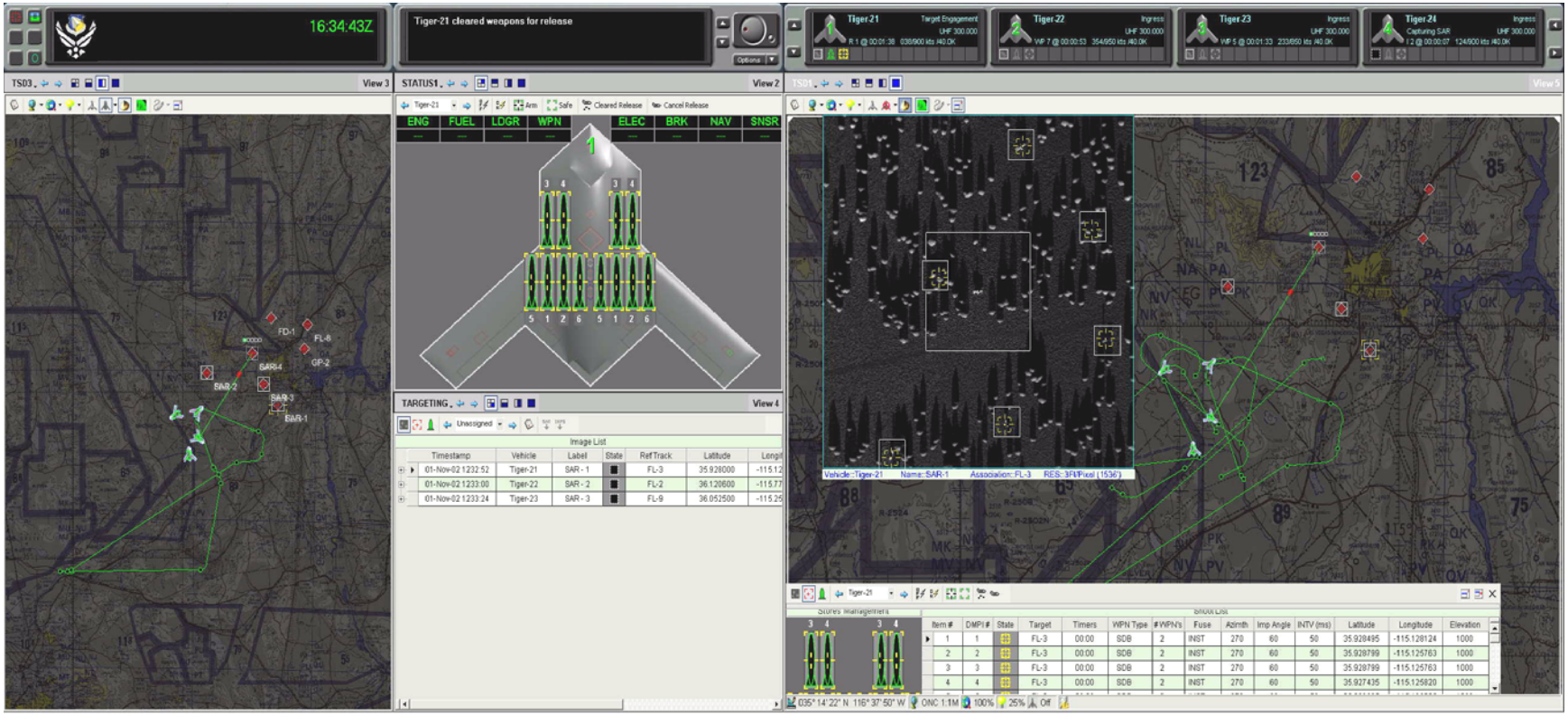

Figure 25. Sample UCAV display showing the SAR processing and target selection.

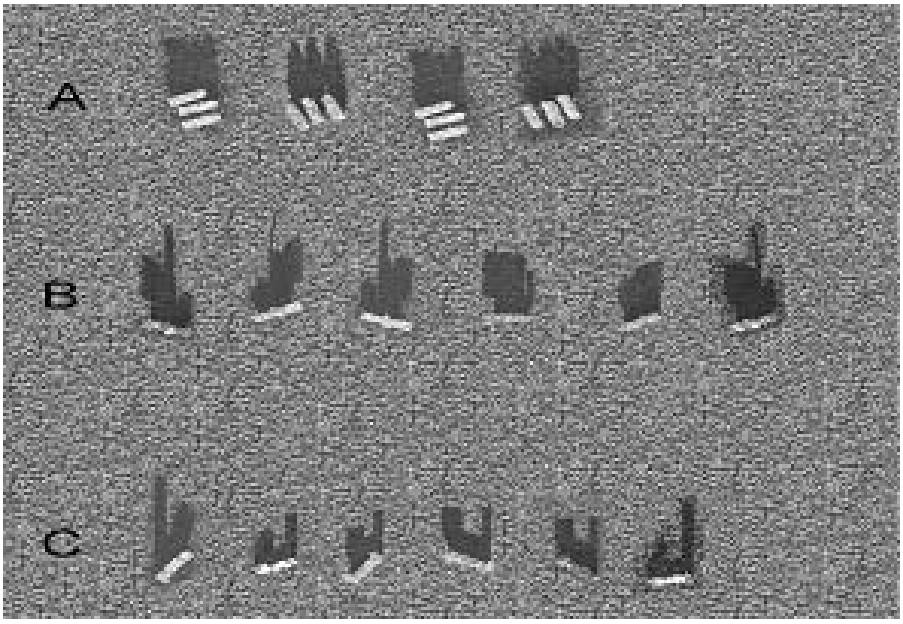

Figure 26. Examples of SAR types: A: Simulated Communication and Command and Control Trailers, B: Simulated SA-10s, and C: Simulated SA-12s.

The OVI task consisted of low and high levels of cognitive workload. The operator had access to twelve weapons per vehicle, and two weapons were allocated to one DMPI in a SAR. Each SAR contained six valid targets as well as distracter targets such as trucks and trees. At the low workload level, the operators were presented with SAR images that contained only six Type C targets; thus, operators could place DMPIs on the targets as soon as they found them in the image.

The high workload level consisted of all target types. The operators were required to search the entire image visually, keeping the location of the targets in spatial working memory before placing the DMPIs. The targets were prioritized according to type: Type C targets were the highest priority, followed by Type B, and finally Type A. The high

workload condition SAR images may contain more than six targets, requiring the operator to remember the priority and track the target type location during an initial visual scan and place the DMPIs on a subsequent scan. For example, a high SAR image may contain three Type C targets, three Type B targets, two Type A targets, and eight distracter targets. The proper response is to place the DMPIs on the three Type C and three Type B targets.

In addition to placing weapons on target, the operators monitored the progress of each vehicle as it flew from waypoint to waypoint. Critical waypoints included a SAR capture waypoint at a predetermined orientation and distance from the target for optimal SAR imaging and a weapons release waypoint, a predetermined point to release the weapons on target for optimal effectiveness. These waypoints were designated during mission planning, and in the case of these experiments, all mission planning was accomplished during the design of the experimental trials to ensure consistency across operators.

After the SAR image was captured at the SAR capture waypoint, the operator downloaded the SAR image to the workstation in approximately sixteen seconds. The operator had to start the SAR image download and place DMPIs on targets in the SAR image, power on weapons, arm the weapons, and clear the weapons for release before the vehicle reached the weapons release waypoint. Omitting any of these steps resulted in partial mission success or complete mission failure. Since each vehicle reached its weapons release points at different times, the operator had to plan the attack to achieve mission completion.

A second task was added to the study to manipulate cognitive workload and to provide information processing based on verbal working memory. A vehicle health task (VHT) was included to enable additional levels of difficulty through a verbal working memory task. The VHT simulated occurrences of system failure. The vehicle systems were categorized by systems type: electrical, mechanical, engine, sensor suite, communication, and system. Each system had two possible types of failure. For example, the electrical system could experience a generator fault or loss of battery power. In that case, the operator must select the correct vehicle from the vehicle drop-down menu (see Figure 27) and then select the appropriate response from the correct system drop-down menu.

During the VHT, two distracter responses were presented in each system drop-down menu. The operator received a text message on the left side of the display directly above the vehicle health task response module (Figure 27) that described the failure and the associated vehicle. For example, if the error text displayed was "Tiger 21 Generator Fault Detected"; the correct response was to select Tiger 21 from the vehicle drop-down menu then to select "Recycle Generators" from the electrical system drop-down menu. A list of possible errors and the correct response pairings is in Appendix A, and a list of all possible responses (including distracter responses) and commands is in Appendix B.

Both the low and high difficulty vehicle tasks were *n*-back memory tasks, which required the operator to retain *n* items in verbal working memory and recall them at a later time (Wickens, 1984). Other call signs were used as distracters for the operator; call signs other than Tiger were to be ignored. The low difficulty level, a 1-back task with one distracter, required the operator to retain one failure-vehicle combination in memory and

ignore a single distracter call sign. The high difficulty workload level was a 4-back memory task with one distracter that required the operator to recall a particular failure for each of the four vehicles. The errors were displayed approximately ten seconds apart. Ten seconds afterwards, the message, "Tiger 21, Execute Solution," appeared telling the operator which vehicle required fault repair. Next, the operator had to recall the error for "Tiger 21" and select the correct repair response. This procedure was repeated several times for the duration of each trial, and the number of cycles depended on the length of the trial.

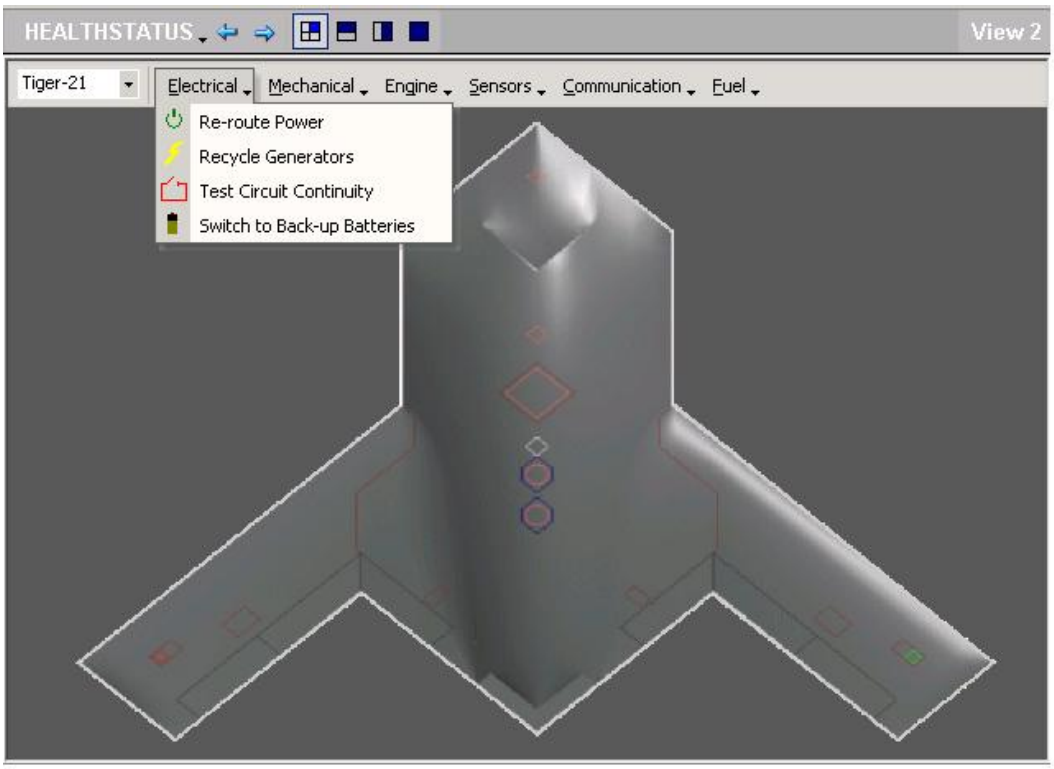

Figure 27. Sample vehicle health task response module, which provides an additional task to drive cognitive load.

3.3 Physiological Measures

The five EEG channels, recorded at sites positioned according to the International 10-20 electrode system (Jasper, 1958), were from electrode sites F7, FZ, T5, PZ, and O2 (see Figure 28). Mastoids were used as reference and ground and all electrode impedances were below 5K ohms. Each EEG channel was corrected for eye movement and blinks using an adaptive filter (He, Wilson, and Russell, 2004) and stored at 200 samples per second. These five sites were selected since previous research (Russell and Gustafson, 2001) showed that they provide the most salient features. Signals from the horizontal and vertical eye and the heart were also collected using a BioRadio 110 manufactured by Cleveland Medical Inc. The signals were transmitted at radio frequencies, eliminating the need to tether operator to amplifiers and to a computer for data collection.

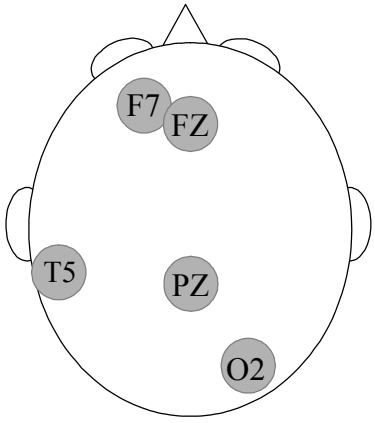

Figure 28. The electrode locations used for operator functional state estimation were determined *a priori* based on previous studies.

Additional measures were collected and evaluated as features for the classifier. Measures collected from an "arousal meter" developed by Clemson University (Schmorrow, 2003) were evaluated as well as electrodermal activity (EDA), electromyographic (EMG) activity, and pupil diameter. Electrodermal activity is the change in electrical activity in the eccrine sweat glands and is influenced by the sympathetic nervous system. Electromyographic activity has been shown to predict arousal accurately (Veldhuizen, Gaillard, and de Vries, 2003) as well as workload (Von Boxtel, Waterink, and Veldhuizen, 1997). Also, changes in pupil diameter can provide estimates of cognitive load (Marshall, 2004).

One-second fast Fourier transforms (FFTs) of the EEG were computed. The power spectra were parsed into frequency bins representing the traditional EEG bands. The frequency ranges of the five traditional bands are delta (~DC-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz), and gamma (31-42 Hz). Time series representations of these bands are shown in Figure 29.

To capture vertical eye movements and eye blinks, electrodes were placed above and below the left eye. Additional electrodes were placed on the right and left side of the head juxtaposed to the right and left eye to collect horizontal eye movements. The vertical and horizontal eye signals were processed the same as the EEG measures, extracting the traditional EEG bands. A blink detection algorithm (Wilson and Russell, 2002) was implemented to compute the time between blinks or interblink interval (IBLI). The algorithm determined blinks by finding the characteristic signal peak caused by eyelid closure followed by the valley caused by the eyelid opening.

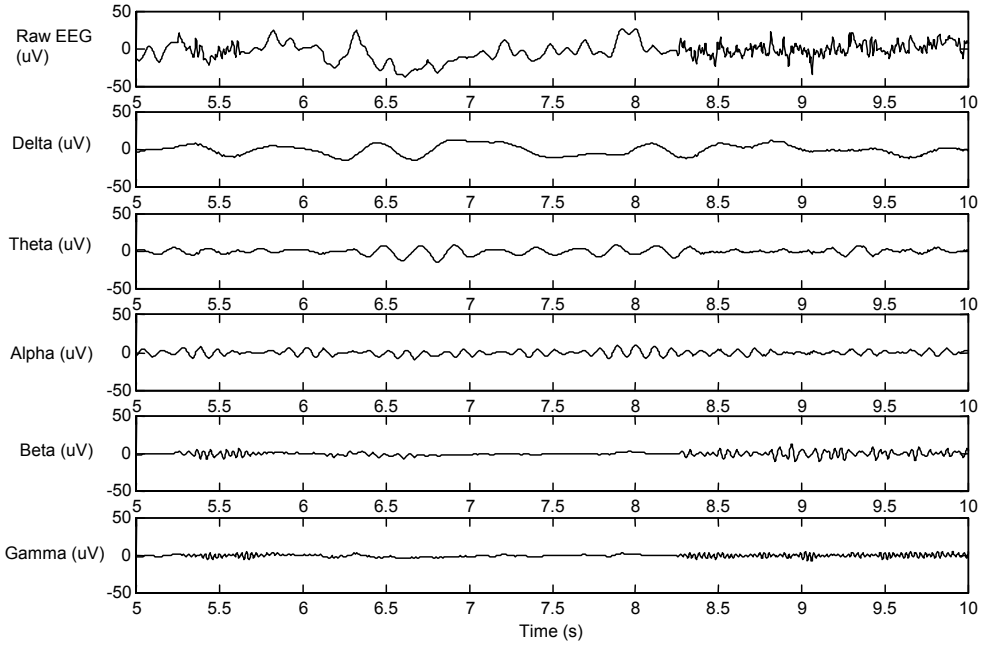

Figure 29. Features were derived from traditional EEG bands, which are bandpass filtered representations of the raw EEG signal.

Additionally, Electrodes were placed at the top of the sternum and the bottom of the rib cage to collect electrocardiographic signals. As stated earlier, these signals were collected with a radiofrequency transmission system and sampled at 200 Hz. A beat detection algorithm (Wilson and Russell, 2002) was implemented to compute the time between the R waves of the heart signal (interbeat interval, IBI), characteristic peaks generated by the closure of the ventricles of the heart.

Pupil area was measured using a head-worn camera-based eye tracking system developed by ISCAN, Inc. This system computed the pupil area and recorded this measure at 60 Hz. Artifacts are caused by eye blinks and are essentially a loss of signal since the camera cannot see the pupil to make a measurement. Blinks were detected using an algorithm that employed a threshold to determine the occurrence and duration of eye

blinks. The signal was then corrected using linear interpolation to recreate the pupil diameter signal.

Electromyograph activity was measured from the corrugator supercilii and frontalis muscles located just above the eyebrow. Developed by TEMEC Instruments, the Vitaport II system recorded signals using bipolar Ag/AgCl electrodes. The signals were lowpass filtered with a cutoff frequency of 32 Hz to eliminate movement artifacts. After filtering, the signals were full-wave linearly rectified and lowpass filtered with a cutoff frequency of 38.4 Hz to smooth the data. The resulting signal was integrated over a 1-second period to produce the final EMG feature.

Additionally, electrodes were placed on the arch of the foot to measure skin conductance or electrodermal activity and were recorded using the Vitaport system. The electrodermal activity was characterized by the tonic or baseline level of the electrodermal signal and recorded at 60 Hz. The arousal meter (Ameter) measures level of arousal based on respiratory sinus arrhythmia - the high frequency component (0.15 – 0.5 Hz) of the heart signal and a known indicator of parasympathetic activity (RTO Human Factors and Medicine Panel Task Group, 2004). The data were collected at 256 Hz and one-second averages of arousal level were computed.

The data were segmented into five-second windows with a four-second overlap as shown in Figure 30. The window and overlap used in this research was determined empirically. Multilayer perceptrons were trained using features processed using a range of window sizes (1 to 20 seconds) and overlap (0 to 19 seconds). The empirical investigation determined that the longer window sizes (5 seconds or more) produced better classification results. However, windows of this length with no overlap could not

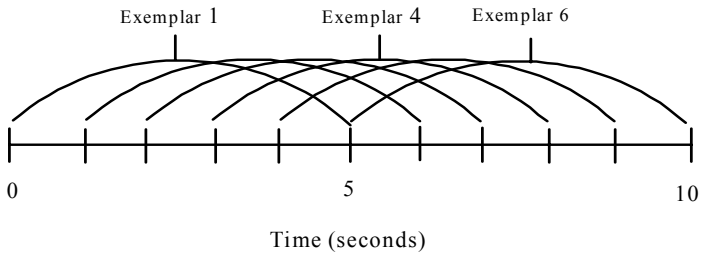

Figure 30. Description of moving window.

provide the update rate required for classifying operator functional state, i.e., classifier outputs would occur every ten or twenty seconds. A one-second update rate was desirable (Wilson, 2003) to enable the adaptive aiding system following a change in operator functional state. The tradeoff between classification accuracy and update rate was considered by varying the window and overlap. A window size of five seconds and an overlap of fours seconds met the one-second update rate and would produce acceptable classification accuracy.

Log power of delta, theta, alpha, beta, and gamma from the five EEG channels and both horizontal and vertical eye channels were used, resulting in 35 EEG features as inputs to the neural network. Three physiologically based features, the interval between eye blinks, heart interbeat intervals, and the Arousal Meter output, were also used as input features, resulting in 38 inputs. These measures were used for all experiments. Additional measures - pupil diameter, integrated muscle activity, and tonic level of

electrodermal activity - were collected off-line and evaluated along with these 38 inputs in a separate study to determine the saliency of these additional measures.

3.4 Performance Measures

The performance data collection consisted of recorded mouse movements, mouse clicks, button presses, VHT prompts and responses, DMPI placements points, vehicle waypoints, heading changes, along with the times these events occurred. The target priorities and the target locations were known for each SAR, so measures such as radial miss distance and time to locate target and place DMPIs were derived. The responses for each of the vehicle health task prompts were recorded as well as the time required to complete the responses.

Coordinates for each DMPI placed within a SAR image were compared to the known locations of each target and distracters within each SAR image. DMPIs were assigned to the nearest target or distracter using a Euclidean distance measure (radial miss distance). Next, signal detection theory was applied to these assignments and hits, misses, false alarms, and correct rejections were computed for each SAR image. Each SAR image had six valid targets and at least six distracter targets. The radial miss distance determined the assignment of operator DMPI placement as either a distracter or a valid target. The number of hits, misses, correct rejections, and false alarms were summed and recorded. For example, consider a SAR containing three type C targets, three type B targets, one type A target and eight distracter vehicles. The operator selected the three C targets, two of the type B targets, and the type A target. The correct targets were the three type C targets and the three type B targets based on the target prioritization

scheme. Thus the signal detection for this SAR was Hits – 5, Misses – 1, Correct Rejections – 8, False Alarms – 1.

Then, mission success was determined for each vehicle. Weapons were not released unless the operator completed a series of tasks - placing DMPIs on targets, powering on the weapons, arming the weapons, and authorizing the release of the weapons - before the UCAV reached the weapons release waypoint. If the weapons were not released by the weapons release waypoint for a particular SAR image, the mission was considered a failure and was scored as a missed weapons release. The number of DMPIs placed on the SAR was also recorded.

The missed weapons release waypoint measures were Bernoulli trials, i.e., a mission was successful or it failed, and a series of Bernoulli trials have a binomial distribution. Additionally, the signal detection measures of hit, miss, correct rejection, and false alarm do not have a normal underlying distribution. Therefore, the Kruskal-Wallis nonparametric test (Rosner, 1995) was used to determine the significance of the differences in the means between the types of aiding and levels of workload. This test was used since the underlying distribution is not normal and the data are ordinal. Pairwise comparisons between the aiding types for each workload level were conducted using the Dunn Procedure (Rosner, 1995). These tests are explained in detail in Appendix C. The results of these tests are reported in Chapter IV and all tests are reported in the form of (N = xxx, z or χ^2 = xxx, p = xxx), where N is the number of data points used for the test, z or χ^2 is the test statistic (z for the Dunn Procedure and χ^2 for the Kruskal-Wallis test), and p is the level of significance of the statistical test.

3.5 Subjective Measures

Subjective data were collected. The NASA Task Load Index (NASA-TLX; Hart and Staveland, 1988) measured the subjective experience of mental workload and scored as a composite of six subscale ratings: mental demand, physical demand, temporal demand, performance, effort, and frustration. Each subscale was scored from low to high and had a numerical range of 0 to 100. The composite score, a weighted combination of these subscales, ranged from 0 to 100, where larger numbers corresponded to greater subjective workload. The operator considered all pairwise comparisons of the subscales and selected one subscale from each comparison as the major contributor to workload from each pair. The weights were the number of times each subscale was considered as the greatest source of workload.

The subjective data have a nearly normal distribution and standard Analysis of Variance (ANOVA) tests can be conducted using these data. Pairwise comparisons between the aiding types for each level of workload were conducted using standard F tests. These test results are reported in Chapter IV and all tests are reported in the form of (n_1 = xxx, n_2 = xxx, F = xxx, p = xxx), where n_1 and n_2 are the number of data points for computing the means for the two groups used in the pairwise comparisons, F is the test statistic, and p is the level of significance of the statistical test.

3.6 Single-Task Experiment

The study was divided into two experiments for data collection. The single-task experiment consisted of trials for four conditions: low Vehicle Health Task (VHT), high VHT, low Operator Vehicle Interface (OVI), and high OVI as defined in Section 3.2. A

trial consisted of one condition. Three randomly presented trials of each condition were conducted to evaluate repeatability.

Several class conditions or gauges were investigated. Spatial working memory, verbal working memory, executive function, global workload, spatial versus verbal working memory, OVI task, and VHT classifiers were developed to determine flexibility of the measures and to provide information on different types of mental demand and task type. The following paragraphs define each gauge and the method of determining the class levels for each gauge.

Working memory is the passive storage of information in memory and is subject to decay (Vidulich, 2004). The items stored in working memory can be maintained through rehearsal but do not stay there unless they receive constant attention. For example, recalling a telephone number after a short period of time requires working memory.

Spatial working memory is maintaining the spatial characteristics of the items in memory. The OVI task contains the spatial working memory component in the single-task study. Because the entire SAR image is not visible at one time, the operator must remember locations of targets in the SAR image and must use both spatial working memory and information from long-term memory to complete the task. The locations of the targets are stored in spatial working memory, and the operator must recall the physical characteristics of the target types to identify them in the SAR image. The operator must also recall the target prioritization schedule from long-term memory.

Classification of spatial working memory levels was represented by the spatial working memory gauge. The spatial working memory gauge consisted of three classes:

no spatial working memory component, low spatial working memory, and high spatial working memory. The no spatial working memory component consisted of data from the low and high VHT trials. The low spatial working memory class consisted of the low OVI trials, and the high spatial working memory class consisted of the high OVI trials.

The VHT drives the verbal working memory in the single-task experiments. After a short time, the operator must recall the vehicle problem as well as which vehicle had the problem. This task requires long-term memory in association with verbal working memory. The operator must also know the appropriate response to a particular problem learned prior to the experiment in the training sessions.

Classification of verbal working memory levels was represented by the verbal working memory gauge. The verbal working memory gauge consisted of no verbal working memory, low verbal working memory, and high verbal working memory. The no verbal working memory class consisted of data from both the low and high OVI tasks. The low verbal working memory class consisted of the low VHT trials, and the high verbal working memory class consisted of the high VHT trials.

A gauge consisting of two classes, verbal working memory and spatial working memory, was examined. This gauge consisted of two classes, verbal working memory and spatial working memory. The verbal working memory class consisted of the low and high VHT trials. The spatial working memory class consisted of the low and high OVI trials.

Another gauge, the executive function, is the high-level processing and planning that accomplishes tasks (Wilson, 2003). This process includes planning and decision making for completing tasks on time and in the correct sequence. This study used a

subscale of the NASA-TLX to determine the levels of executive function among the tasks. The mental demand subscale determined the executive function levels (Vidulich, 2004) using the Tukey-Kramer Honestly Significant Difference test. The number and grouping of the levels were accomplished after all subjects completed all the trials. The mental demand distinguished three levels or classes of executive function: low, medium, and high. The results leading to these classes are discussed in the Results and Analysis section. The low executive function class consisted of the low VHT and low OVI trials, the medium executive function class consisted of the high OVI trials, and the high VHT trials provided the data for the high executive function class.

This study also used global workload to measure the overall workload state. The NASA-TLX composite score determined the levels of global workload among the tasks using a difficulty (Low, High) by working memory (verbal, spatial) analysis of variance (ANOVA). The composite TLX distinguished low and high global workload levels that were determined *a posteriori* after all operators completed the experiment. The results of the analysis leading to this gauge are discussed in the Results and Analysis section. The low global workload class consisted of the low VHT and low OVI trials, and the high global workload class consisted of the high VHT and high OVI trials.

The OVI and VHT gauges were based solely on the respective trials and task conditions. The low VHT class consisted of the low VHT trials, and the high VHT class consisted of the high VHT trials. The OVI trials were separated into the cruise component and the SAR image processing component. The cruise component is the portion of the trial when the operator is not processing a SAR image and mainly

consisted of the ingress to the target portion of the trial. The OVI class condition consisted of three classes: cruise, low SAR, and high SAR.

All combinations of the three trials were used as training and as test sets to train and evaluate the artificial neural networks for each of the gauges (Table 3). Two trials were used as training and one trial was used for testing. A feedforward backpropagation artificial neural network was trained for each gauge. The architecture of the neural network consisted of three layers of fully connected neurons with logistic sigmoid activation functions. The hidden layer consisted of 43 neurons. This training resulted in three artificial neural networks for each subject and each gauge using different trials as training data. For example, group 1 and 2 were used for training the artificial neural network and group 3 was used for testing. A distinct artificial neural network was trained for each of the seven gauges.

Table 3. Trial grouping for the single-task experiment

		Trial Type			
Group	1	Low VHT # 1	High VHT # 1	Low OVI # 1	High OVI # 1
	2	Low VHT # 2	High VHT # 2	Low OVI # 2	High OVI # 2
	3	Low VHT # 3	High VHT # 3	Low OVI # 3	High OVI # 3

3.7 Dual-Task Experiment

The second experiment measured the real-time classification accuracy and the effects of adaptive aiding on operator performance by using the following trials: training, classifier performance, adaptive aided, and randomly aided. The latter three trials were presented randomly.

The dual-task experiments consisted of simultaneously combining the OVI task and VHT to form a complex operational task environment. Both low and high OVI trials were conducted for each of the adaptive aiding trials. Only the high VHT condition was presented with the low and high OVI tasks. Eliminating the low VHT for the dual-task experiments made the analysis less complex and removed confounded conditions.

The first trials for each subject were ANN training trials. Five training trials were conducted - three high conditions and two low conditions for each subject. Two low non-aided dual-task trials and two high non-aided dual-task trials were conducted to evaluate classifier performance. One each of low and high dual -ask trials were conducted for adaptive aided trials, and one each of low and high OVI trials were run for the randomly aided trials.

Furthermore, the ANN training trials were segmented into low and high cognitive load. The cruise segments and processing of the low SAR segments were combined to represent the low cognitive load state. The high cognitive load state consisted of the high SAR segments.

The classifier performance trials were used to evaluate the classifier performance and as baseline trials for comparison to the aided and randomly aided trial types. The

classifier performance trials were also compared to the ANN trials used to train the artificial neural network to ensure consistency across trials.

The aided trials consisted of adaptive aiding triggered by the operator functional state as determined by the artificial neural network classifier. The adaptive aiding consisted of reducing the speed of a vehicle when the ANN detected a high operator functional state. The vehicle the operator was attending decreased speed by 25 percent to allow the operator more time to accomplish the current task. When the ANN detected the operator state had reverted to a nominal workload level, the UCAV continued at its previous speed.

Adaptive aiding was applied randomly during the randomly aided trials. The purpose of the randomly aided trials was to determine the necessity for adaptive aiding; they answer the 'So what?' question. If the randomly aided trials improve operator performance in the same manner as aiding the operator based on cognitive state, either the experimental design is flawed or the aiding should be applied during the entire trial. The total time that the operators were in the high cognitive load state during the OVI trials was partitioned into random starting points throughout the randomly aided trial. For example, during the operator's high aided trial, the ANN detected the high state in six intervals during the trial. The six intervals have different interval lengths. Twice the interval was 15 seconds, once the interval was 30 seconds, and three times the interval was 10 seconds. The randomly aided trial would also have six aided intervals with the same duration as during the aided trial: however, the intervals would occur randomly throughout the randomly aided trial.

3.8 Subjects

Seven subjects, four males and three females, were paid for their voluntary participation in this study. All were right-handed and had normal or normal corrected vision. All subjects signed informed consent documentation approved by the AFRL Human Use Committee.

All subjects were also trained to stabilize performance and eliminate learning effects. Such training usually required several trials over two or three days, depending on subject ability. Stable performance consisted of repeated, reliable performance over several trials until the subject developed a consistent strategy for completing the required task. One subject could not perform the dual task and was removed from the study.

The collection of data in human-subject experiments is difficult and time consuming. Subjects require training in order to perform the experiments with consistent results based on the manipulations in the study design. This is required to preclude effects of learning. Large quantities of performance data are not feasible in studies consisting of human-subject experiments. For example, in the case of the dual task experiment, to collect the 32 samples of the missed weapons release measure for a single subject required approximately 24 hours of training and 4 hours of data collection.

3.9 Classifier Evaluation

The voluminous data collected from human studies requires methods suited for high-capacity data. Human studies usually generate megabytes or even gigabytes of data from each subject, necessitating classifiers that can operate on very large data sets and still learn appropriate models quickly for real-time applications. Classifiers in this study are of three classes: DA, ANNs, and SVMs. The discriminant analysis techniques used

three discriminant functions: linear, quadratic, and logistic. The ANN used is a feedforward multilayer perceptron with backpropagation training. The architecture of the neural network consisted of three layers of fully connected neurons with logistic sigmoid activation functions. The input layer consisted of 43 neurons that corresponded to the number of input features. The hidden layer consisted of 43 neurons, and the output layer consisted of 2, 3, or 4 neurons, depending on the cognitive model being developed. The support vector machines examined three specific inner product kernels: linear, polynomial, and radial basis functions.

The data from the single-task experiment were used to compare the performance of the classification algorithms. The data were processed by the same procedures used in the single-task experiment. The same training and test data were presented to each of the classifiers to allow for direct comparison for each of the cognitive gauges: spatial working memory, verbal working memory, executive function, global workload, spatial versus verbal working memory, OVI task, and VHT.

3.10 Section Summary

This section described the methods of this research. Methods for collecting psychophysiological signals, performance measures, and subjective ratings were discussed. Processing raw signals into useable features for classification was described, cognitive gauges were defined, and classifier comparison methods used in this research were discussed. Finally, the study design for two experiments was described and the rationale for conducting these experiments was presented.

IV. Results and Analysis

4.1 Single Task-Analysis

4.1.1 Subjective Workload Analysis

The subjective workload data were analyzed with a difficulty (low, high) by working memory (verbal, spatial) analysis of variance (ANOVA). As shown in Figure 31, difficulty (low and high workload) manipulation had a significant effect on the NASA-TLX workload scores ($n_1 = 36$, $n_2 = 36$, $F = 56.3$, $p = 0.0007$), but no significant effects of working memory (verbal and spatial) were detected ($n_1 = 36$, $n_2 = 36$, $F = 0.443$, $p = 0.508$). These results indicate that the tasks have different levels of workload based on subjective measures, as desired.

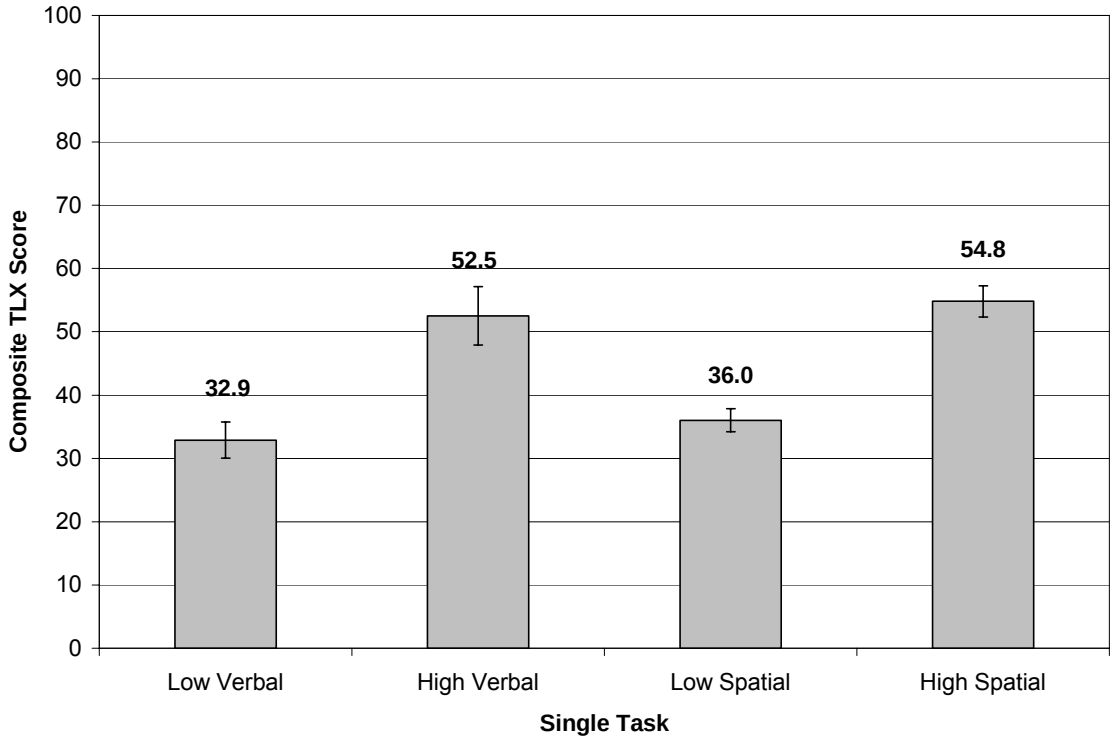

Figure 31. Group means of composite NASA-TLX rating with standard error of the mean for single-task analysis show good separation between low and high cognitive load.

As described in the methodology section, the NASA-TLX mental demand subscale was analyzed to determine the levels of executive function in training the neural network for the executive function gauge. The Tukey-Kramer Honestly Significant Difference (HSD) (Sall, Lehman, and Creighton, 2001) was used to determine differences between the conditions. This test was selected since it is more conservative: the least significant difference intervals of the Tukey-Kramer HSD are larger than the Student's t intervals. These tests also adjust the probability or level of significance for multiple comparisons. The low executive function consisted of the low verbal and low spatial working memory, the medium executive function difficulty consisted of the high spatial working memory task, and the high executive function consisted of the high verbal working memory task. A visual representation is shown in Figure 32. Table 4 shows the results of the Tukey-Kramer HSD analysis, where positive values show pairs of means that are significantly different. For example, the low verbal trials were significantly different than both the high verbal and high spatial trials but not the low spatial trials. Higher values indicate a higher level of significance.

In summary, the low and high conditions for each of the working memory tasks show significant differences, indicating the levels of workload are distinct to the operator and reinforcing the study design. Also, the NASA-TLX mental demand subscale distinguishes three levels of executive function in the trials.

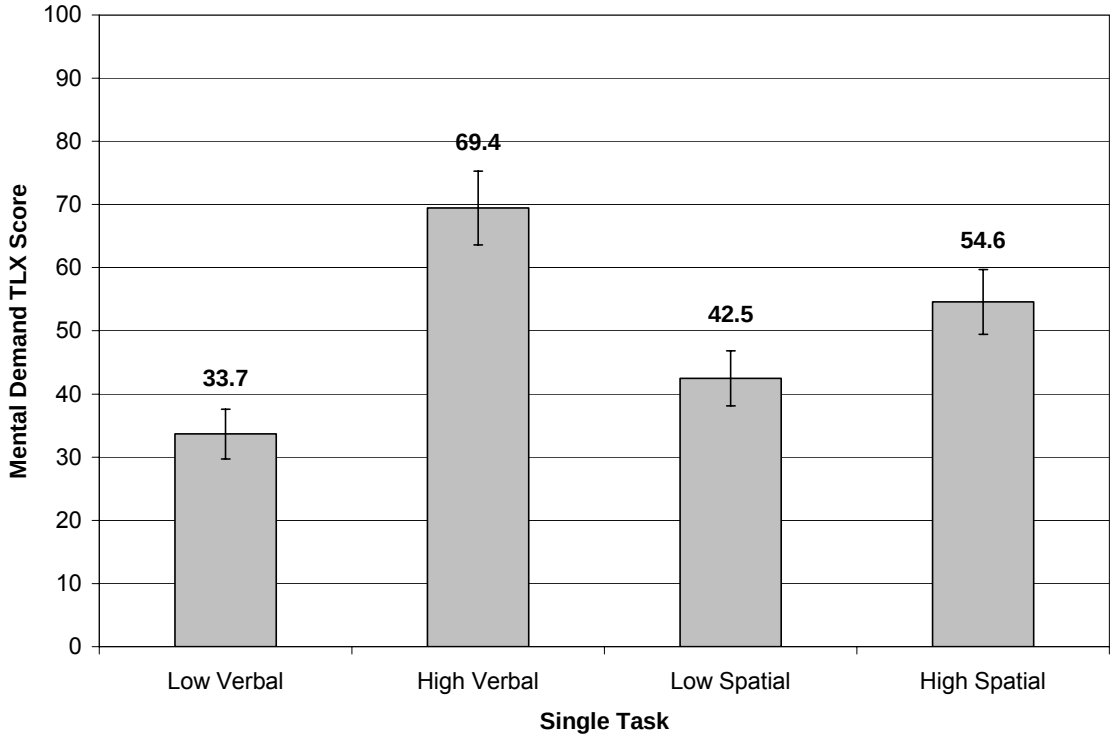

Figure 32. Group means of mental demand TLX subscale with standard error of the mean for single-task analysis were used to determine the class groups for the executive function.

Table 4. Tukey-Kramer HSD comparisons of NASA-TLX mental demand

	High Verbal	High Spatial	Low Spatial	Low Verbal
High Verbal	-18.08	3.05*	8.86*	15.65*
High Spatial	3.05*	-18.08	6.17*	0.62*
Low Spatial	8.86*	6.17*	-18.08	-11.29
Low Verbal	15.65*	0.62*	-11.29	-18.08

* Significant effect at $p < 0.05$

4.1.2 Operator Performance Analysis

4.1.2.1 OVI Task Performance

The operator performance data for the OVI task data were analyzed for the effect of difficulty using a (low, hHigh) Kruskal-Wallis nonparametric test (see Appendix C for

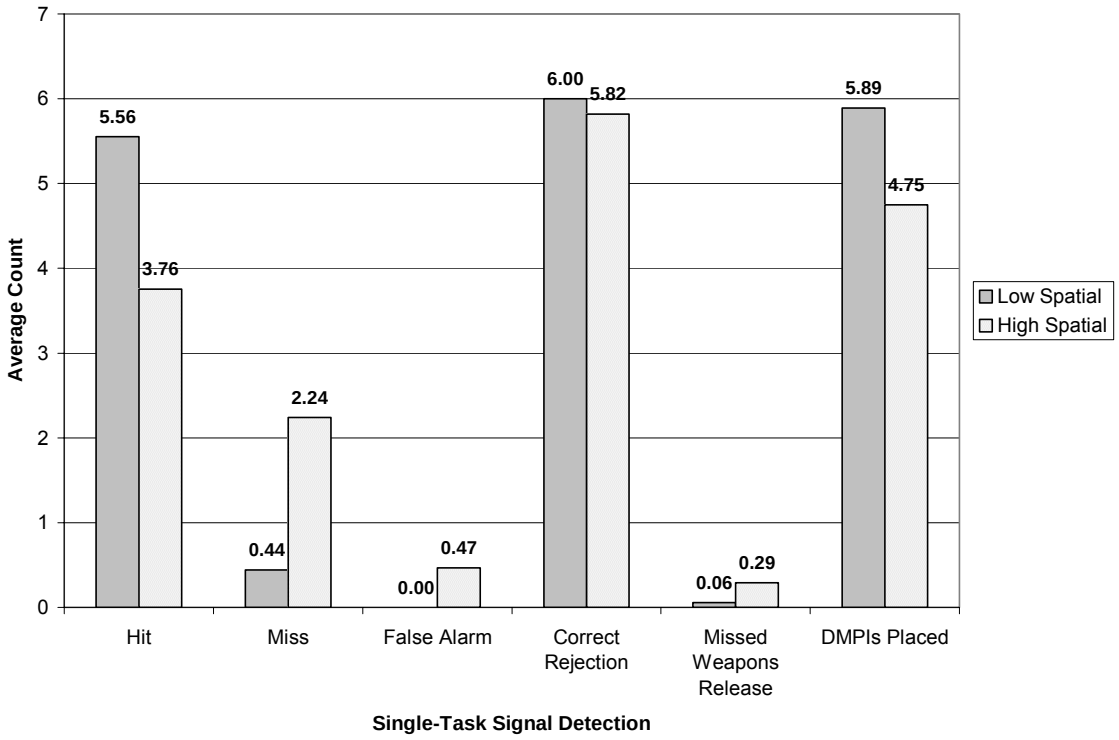

Figure 33. Group means of signal detection for OVI Task performance for single-task analysis.

a complete discussion of the Kruskal-Wallis test). Each performance measure was compared separately and is displayed in Figure 33 for comparison. All measures showed a significant difference between the low and high spatial working memory for OVI task performance. Using signal detection theory, the performance measures developed were Hit ($N = 144$, $\chi^2 = 32.9$, $p < 0.0001$), Miss ($N = 144$, $\chi^2 = 32.9$, $p < 0.0001$), False Alarm ($N = 144$, $\chi^2 = 18.9$, $p < 0.0001$), and Correct Rejection ($N = 144$, $\chi^2 = 18.9$, $p < 0.0001$). The mission success measures were Missed Weapons Release ($N = 144$, $\chi^2 = 13.9$, $p = 0.0002$) and Number of DMPIs Placed ($N = 144$, $\chi^2 = 19.7$, $p < 0.0001$). These results show that the two levels of workload result in significant differences in operator performance, as desired.

In summary, the data show significant differences in operator performance in both the measures derived from signal detection theory and those related to mission success. The operators missed more targets and selected more wrong targets in the high workload condition than in the low workload condition. In the high condition, the operators missed 30% of their weapons release points, resulting in partial mission failure. These analyses are the expected results for the single-task experiment.

4.1.2.2 VHT Performance

The operator performance data for the VHT data were analyzed with a difficulty (low, high) ANOVA. Correct and incorrect responses show a significant effect of verbal demand ($n_1 = 18$, $n_2 = 18$, $F = 7.56$, $p = 0.0105$) in difference in the means: 84% for the high VHT and 49% for the low VHT.

The high verbal condition had a 35% decrease in correct responses, indicating operator difficulty in recalling the problems associated with a particular vehicle. The results indicate that the levels of difficulty are significantly different, as expected.

4.1.3 Cognitive State Classification

Multilayer perceptrons using backpropagation training were trained and tested for each cognitive gauge as described in Section 3.5. Four ANNs were trained for each subject and each cognitive gauge, resulting in a total of 168 trained ANNs. Overall classification results for each gauge are shown in Figure 34. All test results are above chance in randomly selecting a class. Confusion matrices were compiled to determine selectivity and specificity and are attached as Appendix F.

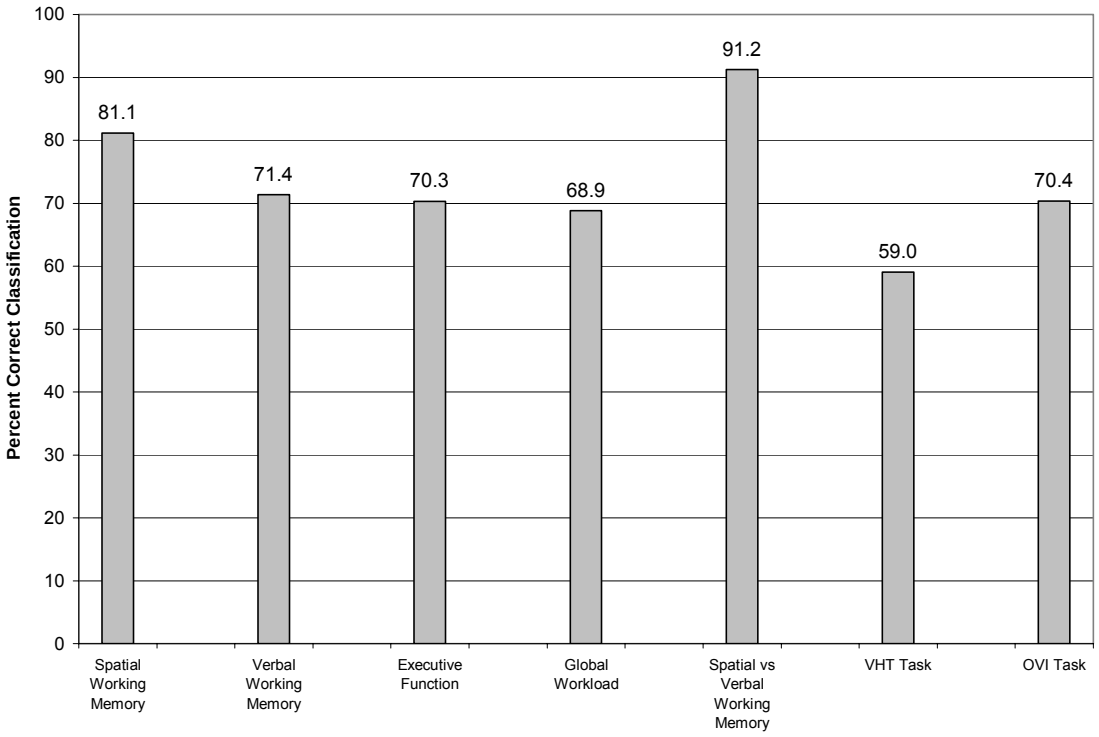

Figure 34. Classification accuracy for the various cognitive gauges.

Overall classification accuracy ranged from 59.0% to 91.2% depending on the cognitive gauge tested. The best results occurred when classifying spatial and verbal working memory, with an overall 91.2% accuracy. The spatial versus verbal working memory classifier showed good specificity between the classes. The verbal working memory was correctly classified in 89.5% of the test data while the spatial working memory data was correctly classified 92.9% across all subjects as shown in Appendix F. These percentages indicate that the psychophysiological measures used in this study can distinguish two information processing cognitive tasks accurately. These results can be used to enhance adaptive aiding by tailoring mitigations based on information context.

The classifiers for the VHT and the verbal working memory gauges did not perform as well as expected. In both cases, the specificity was poor. Futhermore, the low

VHT and the low verbal working memory as well as the high VHT and the high verbal working memory performed similarly. The only difference in the data presented to the gauges was an additional class of verbal working memory. The verbal working memory also contained a class of no verbal working memory. This lack of specificity in the classifiers may indicate that the psychophysiological measures used in this study do not allow differentiation between levels of verbal working memory.

The classifier for the executive function gauge also has poor specificity in the medium and high executive function classes, possibly meaning the NASA-TLX mental demand subscale is not a good indicator of executive function. Other possible explanations for the poor classifier performance are the location of the EEG electrodes or these tasks are not representative of executive cognitive function. The former is not likely; executive function is associated with the frontal lobe of the brain and two of the EEG electrodes measured frontal lobe activity. The latter is a possible source of the problem. Executive function is associated with high level planning, and both the spatial OVI task and the verbal VHT have little planning involved in their execution.

4.2 Dual-Task Analysis

4.2.1 Subjective Workload Analysis

The subjective workload data were analyzed with a difficulty (low, high) by aiding type (no-aiding – training, no-aiding, aiding, and random aiding) ANOVA. As shown in Figure 35, the difficulty manipulation had a significant effect on the NASA-TLX workload scores ($n_1 = 42$, $n_2 = 42$, $F = 22.1$, $p = 0.0053$).

Contrast comparisons for all groups were made for the low and high workload conditions pairwise by aiding type. Table 5 contains the contrast comparisons for the low

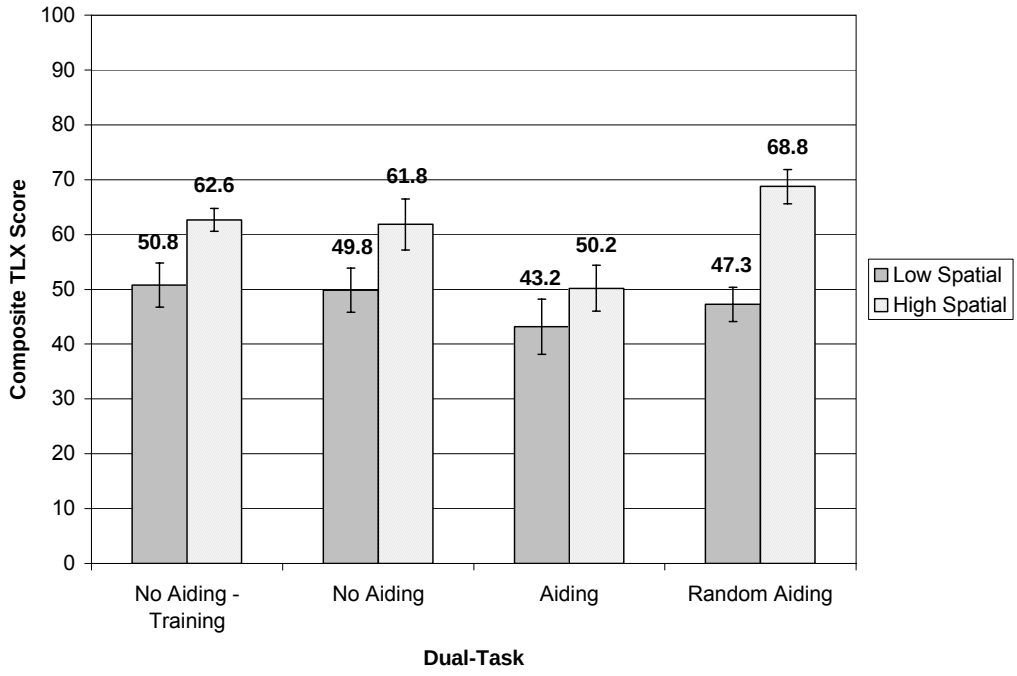

Figure 35. Group means composite NASA-TLX rating with standard error of the mean for dual-task analysis.

workload trials and Table 6 displays the contrast comparisons for the high workload trials. No significant effects of aiding type were found in the low workload condition. These results indicate the operators perceived no reduction in cognitive workload when adaptive aiding was presented.

Table 5: Contrast Comparisons for Low Workload by Aiding Type

	Training	No-aiding	Aiding	Random Aiding
Training		F = 0.0335 p = 0.856		
No-aiding	F = 0.0335 p = 0.856		F = 2.670 p = 0.123	F = 0.399 p = 0.537
Aiding		F = 2.670 p = 0.123		F = 0.754 p = 0.399
Random Aiding		F = 0.399 p = 0.537	F = 0.754 p = 0.399	

* Significant effect at $p < 0.05$

Table 6: Contrast Comparisons for High Workload by Aiding Type

	Training	No-aiding	Aiding	Random Aiding
Training		F = 0.0295 p = 0.865		
No-aiding	F = 0.0295 p = 0.865		F = 8.22 * p = 0.0117	F = 2.91 p = 0.109
Aiding		F = 8.22 * p = 0.0117		F = 6.401 * p = 0.0156
Random Aiding		F = 2.91 p = 0.109	F = 6.401 * p = 0.0156	

* Significant effect at $p < 0.05$

Contrast comparisons were made for the low and high workload conditions between aiding type training and no-aiding to verify that these conditions were the same since the trials were similar. No significant effect existed for both the low and high workload conditions between no-aiding and training aiding types. Since these trials were similar and the operators perceived no differences in cognitive workload in both the low and high workload trials, this result indicates that these conditions were the same, as should be the case since no adaptive aiding was presented in any of the training and no-aiding trials.

A significant effect of aiding type occurred for the contrast comparison of no-aiding and aiding for the high condition. A significant effect was noted for the comparison between aiding and random aiding but not between no-aiding and random aiding. These results indicate that applying adaptive aiding to the OVI task yields a significant decrease in subjective operator workload. The operator does not perceive a significant decrease in workload when adaptive aiding is presented randomly. In fact, the operator does not perceive any difference between random aiding and no aiding. These results indicate that applying adaptive aiding to the OVI task must be presented at the

appropriate times based on operator functional state. Randomly aiding the operator does not decrease the operator's perceived workload. The adaptive aiding provided during the OVI tasks reduces the perceived workload of the operators.

4.2.2 Operator Performance Analysis

4.2.2.1 OVI Task Performance

The operator OVI task performance data were compared using a 1-variable difficulty (low, high) ANOVA with the data collapsed across aiding type. Results show a significant effect of workload across all performance measures. The performance measures were Hit ($N = 192$, $\chi^2 = 25.7$, $p < 0.0001$), Miss ($N = 192$, $\chi^2 = 25.7$, $p < 0.0001$), False Alarm ($N = 192$, $\chi^2 = 8.17$, $p = 0.0043$), and Correct Rejection ($N = 192$, $\chi^2 = 8.25$, $p = 0.0041$). The mission success measures were Missed Weapons Release ($N = 192$, $\chi^2 = 16.5$, $p < 0.0001$) and Number of DMPIs Placed ($N = 192$, $\chi^2 = 24.1$, $p < 0.0001$). Operator performance is poorer for the high workload trials regardless of aiding type - an expected result since the study was designed to include two distinct levels of workload.

OVI performance measures include the signal detection theory measures, hit, miss, false alarm, and correct rejection, and mission performance measures of missed weapons release and number of DMPIs placed per SAR image. The frequencies of these measures are tabulated for each aiding type and workload level in Table 7. These measures were analyzed using a Kruskal-Wallis test. Results are displayed in Figure 36. Contrast comparisons for all groups were made for the low and high workload conditions pairwise by aiding type and by performance measure based on the z score of the rank

Table 7: Frequency of Hit, Miss, False Alarms, and Correct Rejection by Aiding Type and Workload Level Using the Raw Data in Appendix G.

	Hit	Miss	False Alarm	Correct Rejection
Low Workload No Aiding	274	14	0	288
Low Workload Aiding	144	0	0	144
Low Workload Random Aiding	143	1	1	143
High Workload No Aiding	197	91	6	285
High Workload Aiding	127	17	7	141
High Workload Random Aiding	97	47	11	140

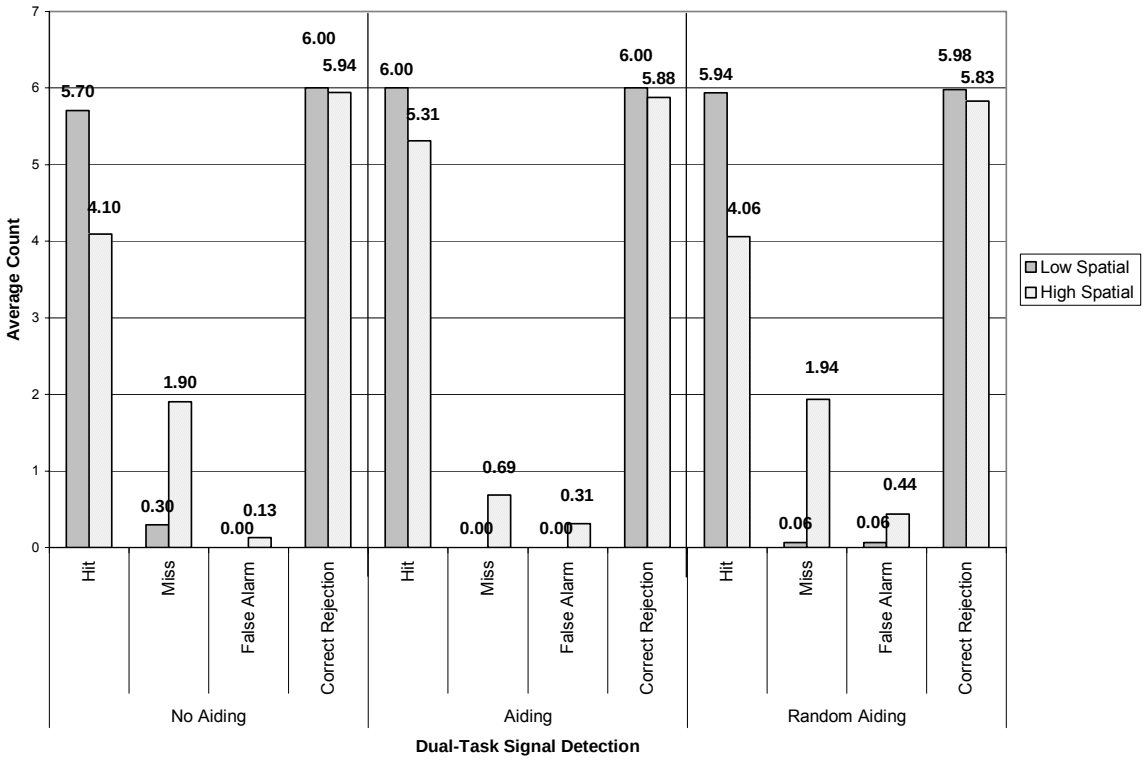

Figure 36. Group means signal detection for dual-task OVI performance analysis.

sums derived during the Kruskal-Wallis analysis using the Dunn Procedure (Rosner, 1995). A complete discussion of the Dunn Procedure appears in Appendix C. Each performance measure is discussed in turn.

Contrast comparisons for all groups were made for the low and high workload conditions pairwise by aiding type for the signal detection performance measure of hits. Table 8 contains the contrast comparisons for the low workload trials and Table 9 displays the comparisons for the high workload trials. No significant effects of aiding type were found in the low workload condition for the hits performance measure. The high workload has significant effects of aiding type between the no-aiding type and the aiding type and a significant effect of aiding type between the aiding and the random aiding conditions. There was no significant effect of aiding type between the no-aiding trials and random aiding type for the high workload condition, demonstrating that randomly aiding the operator does not improve performance. The aiding must be presented at the appropriate time based on operator functional state.

Table 8: Contrast Comparison for Hits During Low Workload

	No-aiding	Aiding	Random Aiding
No-aiding		$z = 0.481$ $p = 0.631$	$z = 0.481$ $p = 0.631$
Aiding	$z = 0.481$ $p = 0.631$		$z = 0.721$ $p = 0.470$
Random Aiding	$z = 0.481$ $p = 0.631$	$z = 0.721$ $p = 0.470$	

* Significant effect at $p < 0.05$

Table 9: Contrast Comparison for Hits During High Workload

	No-aiding	Aiding	Random Aiding
No-aiding		z = 3.32 * p =0.0009	z = 1.58 p = 0.114
Aiding	z = 3.32 * p =0.0009		z = 3.68 * p = 0.0002
Random Aiding	z = 1.58 p = 0.114	z = 3.68 * p = 0.0002	

* Significant effect at $p < 0.05$

Table 10 contains contrast comparisons for the low workload trials and Table 11 displays the contrast comparisons for the high workload trials for the miss performance measure. No significant effects of aiding types were found in the low workload condition. The high workload has significant effects of aiding type between the no-aiding and the aiding and no significant effect of aiding between the no-aiding trials and random aiding for the high workload condition. In addition, a significant effect of aiding type was found between the random aiding and the aiding conditions.

Table 10: Contrast Comparison for Misses During Low Workload

	No-aiding	Aiding	Random Aiding
No-aiding		z = 0.481 p = 0.631	z = 0.481 p = 0.631
Aiding	z = 0.481 p = 0.631		z = 0.721 p = 0.471
Random Aiding	z = 0.481 p = 0.631	z = 0.721 p = 0.471	

* Significant effect at $p < 0.05$

Table 11: Contrast Comparison for Misses During High Workload

	No-aiding	Aiding	Random Aiding
No-aiding		z = 3.32 * p =0.0009	z = 1.58 p = 0.114
Aiding	z = 3.32 * p =0.0009		z = 3.68 * p = 0.0002
Random Aiding	z = 1.58 p = 0.114	z = 3.68 * p = 0.0002	

* Significant effect at $p < 0.05$

The false alarm performance measure contrast comparisons are displayed in the following tables. Table 12 contains the comparisons for the low workload trials and Table 13 displays the comparisons for the high workload trials for the false alarm performance measure. No significant effects of aiding type were found in the low workload condition for this measure. The high workload has significant effects of aiding type between the no-

Table 12: Contrast Comparison for False Alarms During Low Workload

	No-aiding	Aiding	Random Aiding
No-aiding		z = 0 p = 1	z = 1.087 p = 0.277
Aiding	z = 0 p = 1		z = 1.087 p = 0.277
Random Aiding	z = 1.087 p = 0.277	z = 1.087 p = 0.277	

* Significant effect at $p < 0.05$

Table 13: Contrast Comparison for False Alarms During High Workload

	No-aiding	Aiding	Random Aiding
No-aiding		z = 1.73 p = 0.0837	z = 2.92 * p = 0.0035
Aiding	z = 1.73 p = 0.0837		z = 0.890 p = 0.374
Random Aiding	z = 2.92 * p = 0.0035	z = 0.890 p = 0.374	

* Significant effect at $p < 0.05$

aiding trials and the random aiding trials. No significant effect of aiding type existed between the no-aiding trials and aiding trials or between the aiding and random aiding trials for the high workload condition.

These results do not agree with the hypotheses established in the study. One would expect the false alarms to be reduced when the operator is being aided at the appropriate times. The cause of this disparity with the study hypothesis could be the power of the tests conducted due to the infrequent occurrence of false alarms.

Contrast comparisons for correct rejection for the dual-task experiment showed the same results as the false alarm condition. Table 14 contains the comparisons for the low workload trials, and Table 15 displays the comparisons for the high workload trials

Table 14: Contrast Comparison for Correct Rejection During Low Workload

	No-aiding	Aiding	Random Aiding
No-aiding		z = 0 p = 1	z = 1.060 p = 0.289
Aiding	z = 0 p = 1		z = 0.795 p = 0.427
Random Aiding	z = 1.060 p = 0.289	z = 0.795 p = 0.427	

* Significant effect at $p < 0.05$

Table 15: Contrast Comparison for Correct Rejection During High Workload

	No-aiding	Aiding	Random Aiding
No-aiding		z = 1.66 p = 0.0964	z = 2.85 * p = 0.0044
Aiding	z = 1.66 p = 0.0964		z = 0.890 p = 0.374
Random Aiding	z = 2.85 * p = 0.0044	z = 0.890 p = 0.374	

* Significant effect at $p < 0.05$

for the correct rejection performance measure. The correct rejection values were normalized to a value of six to compare results. Unlike the target data, which is fixed at six targets possible, the numbers of distracter targets vary by SAR image. No significant effects of aiding type were found in the low workload condition for the correct rejection performance measure. The high workload has significant effects of aiding between the no-aiding and the random aiding. No significant effect of aiding type existed between the no-aiding trials and aiding trials or between the aiding and random aiding trials for the high workload condition. The results show the targets correctly rejected were not significantly decreased when the operator was adaptively aided.

Mission performance measures include a vehicle missing the weapons release waypoint and the number of DMPIs assigned to each SAR image (Tables 16 and 17). The missed weapons release point is a mission failure. This performance measure is computed as a ratio of the missed weapons release waypoint and the total number of weapons release waypoints in a trial and are presented in Figures 37 and 38.

Table 16: Frequency of Mission Success by Aiding Type and Workload Level Computed Using the Raw Data in Appendix H.

	Success	Failure	N (Total Count)
Low Workload No Aiding	46	2	48
Low Workload Aiding	24	0	24
Low Workload Random Aiding	24	0	24
High Workload No Aiding	36	12	48
High Workload Aiding	22	2	24
High Workload Random Aiding	18	6	24
Total	170	22	192

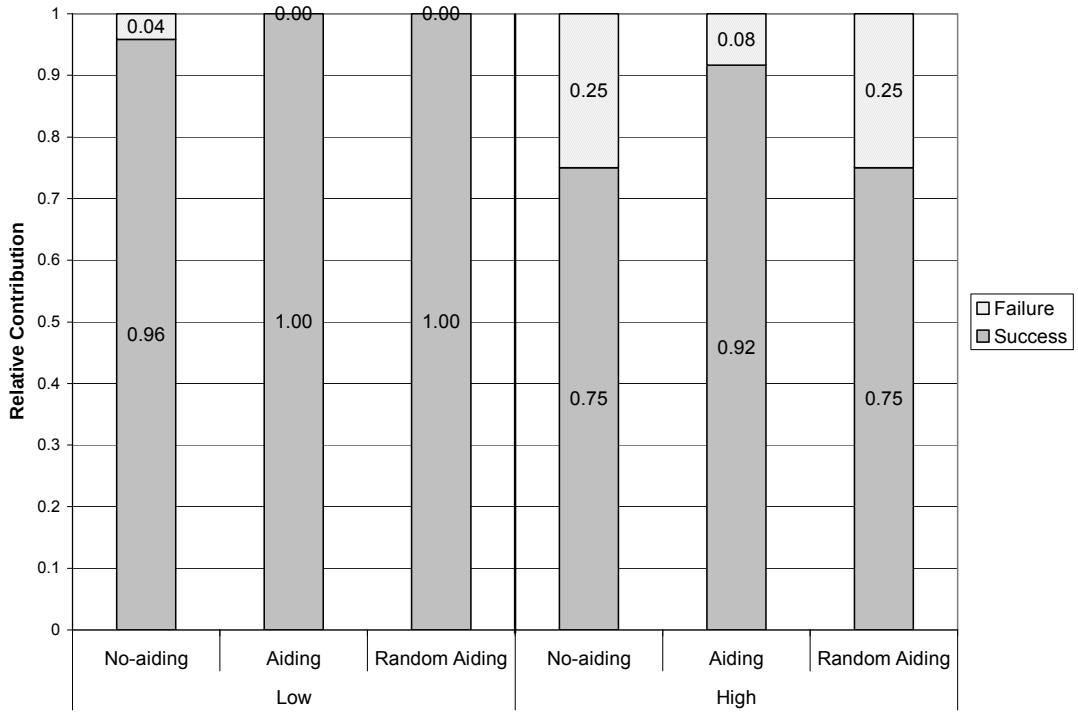

Figure 37. Occurrences of successful and unsuccessful completion mission requirements for the weapons release waypoints.

Table 17: Frequency of Placed DMPIs by Aiding Type and Workload Level Computed Using the Raw Data in Appendix H.

	0	1	2	3	4	5	6
Low Workload No Aiding	0	0	0	1	1	0	48
Low Workload Aiding	0	0	0	0	0	0	24
Low Workload Random Aiding	0	0	0	0	0	0	24
High Workload No Aiding	3	0	5	3	1	4	32
High Workload Aiding	0	0	0	1	1	1	21
High Workload Random Aiding	2	2	2	0	0	1	17

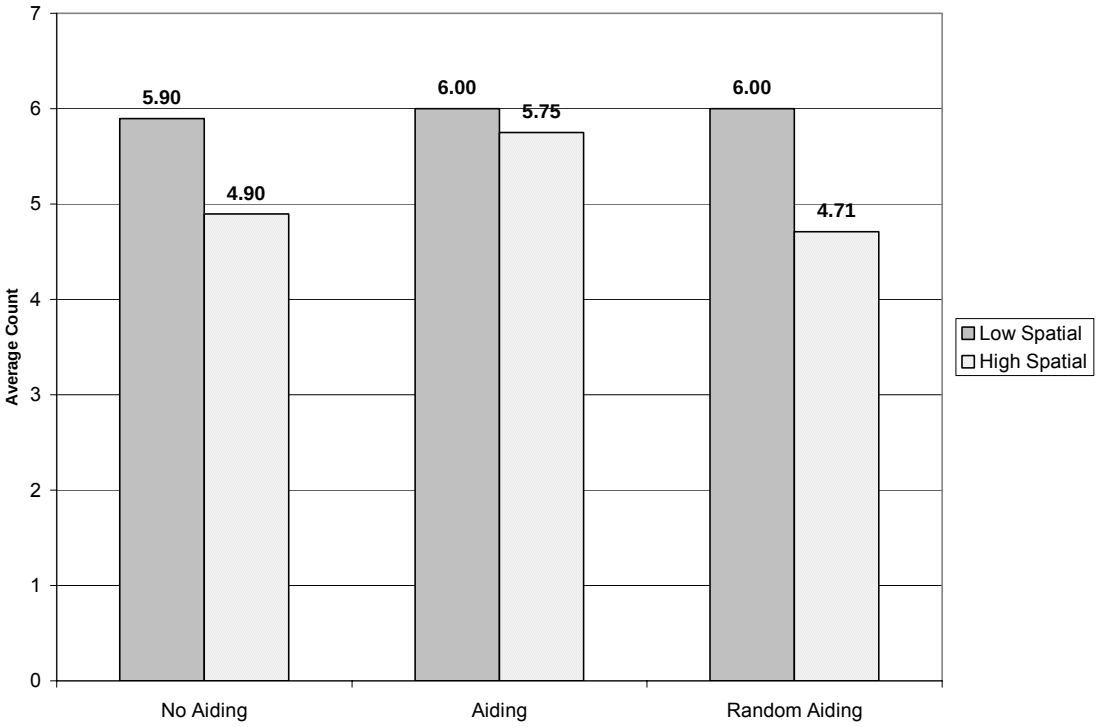

Figure 38. Average count of numbers of placed DMPIs for each aiding condition and workload level.

Table 18 contains comparisons for the low workload trials and Table 19 displays the comparisons for the high workload trials for the missed weapons release performance measure. No significant effects of aiding type were found in the low workload condition. The high workload has significant effects of aiding type between the no-aiding trials and the aiding trials. There was a significant effect of aiding type between the aiding trials and random aiding for the high workload condition but none for the comparison of no-aiding and random aiding under the high cognitive workload condition. The mission effectiveness was improved with the implementation of adaptive aiding. However, if the aiding was not presented appropriately as was presented during the random aided trial, aiding did not improve mission effectiveness. These results suggest the aiding must be presented at the appropriate time to improve mission effectiveness.

Table 18: Contrast Comparison for Missed Weapons Release During Low Workload

	No-aiding	Aiding	Random Aiding
No-aiding		z = 1.15 p = 0.249	z = 1.15 p = 0.249
Aiding	z = 1.15 p = 0.249		z = 0 p = 1
Random Aiding	z = 1.15 p = 0.249	z = 0 p = 1	

* Significant effect at $p < 0.05$

Table 19: Contrast Comparison for Missed Weapons Release During High Workload

	No-aiding	Aiding	Random Aiding
No-aiding		z = 4.61 * p < 0.0001	z = 0 p = 1
Aiding	z = 4.61 * p < 0.0001		z = 3.46 * p = 0.0006
Random Aiding	z = 0 p = 1	z = 3.46 * p = 0.0006	

* Significant effect at $p < 0.05$

The operators missed the weapons release points on average 25% of the time for both the no-aiding condition and the random aiding condition (see figure 37). However, aiding the operator at the appropriate time reduced the missed weapons release to 8% of the missions on average, which is a 67% ± 3% improvement in mission effectiveness. The 3% error was computed based on the loose assumption that the population standard deviation for the failures both the aided and no-aided high workload trials (see Table 16) is one trial. Then, the corresponding standard deviations in the mean numbers of no-aided and aided trials are $s = \frac{1}{\sqrt{48}}$ and $t = \frac{1}{\sqrt{24}}$, respectively. Assuming the failures are independent then the variance in the improvement (67%) in missed weapons release waypoints is $v = \left(\frac{2}{12}\right)^2 s^2 + \left(\frac{2*2}{12^2}\right) t^2 = 0.03^2$. Thus, for the specified population

variance and independence assumptions, the improvement in missed weapons release waypoints is 67% ± 3%. Similar computations can be made for all of the performance measures.

Table 20 contains comparisons for the low workload trials and Table 21 displays the comparisons for the high workload trials for the number of DMPIs placed mission performance measure. No significant effects of aiding type were found in the low workload condition for the mission performance measure. The high workload has significant effects of aiding type between the random aiding and the aiding trials. There was also a significant effect of aiding type between the no-aiding trials and aiding trials for the high workload condition.

Table 20: Contrast Comparison for Number of DMPIs Placed During Low Workload

	No-aiding	Aiding	Random Aiding
No-aiding		z = 1.098 p = 0.272	z = 1.098 p = 0.272
Aiding	z = 1.098 p = 0.272		z = 0 p = 1
Random Aiding	z = 1.098 p = 0.272	z = 0 p = 1	

* Significant effect at $p < 0.05$

Table 21: Contrast Comparison for Number of DMPIs Placed During High Workload

	No-aiding	Aiding	Random Aiding
No-aiding		z = 5.96 * p < 0.0001	z = 0.696 p = 0.487
Aiding	z = 5.96 * p < 0.0001		z = 3.95 * p < 0.0001
Random Aiding	z = 0.696 p = 0.487	z = 3.95 * p < 0.0001	

* Significant effect at $p < 0.05$

The number of DMPIs placed for the aided task increased from 4.9 to 5.8 over the number of DMPIs places for the no-aiding task - an increase of almost an additional target per SAR image, resulting in an additional four targets destroyed per vehicle for an entire mission. The number of DMPIs placed in the randomly aided trials decreased relative to the number of DMPIs placed in the aided trials.

4.2.2.2 VHT Performance

The performance of the vehicle health task degraded considerably from the performance in the single-task experiment. In the single-task experiment, the operators responded correctly to about half of the prompts. In the dual-task experiment, the operators had a 7 to 18% reduction in correct responses in the low trials and an 11 to 18% reduction in correct response in the high trials. Results are shown in Figure 39.

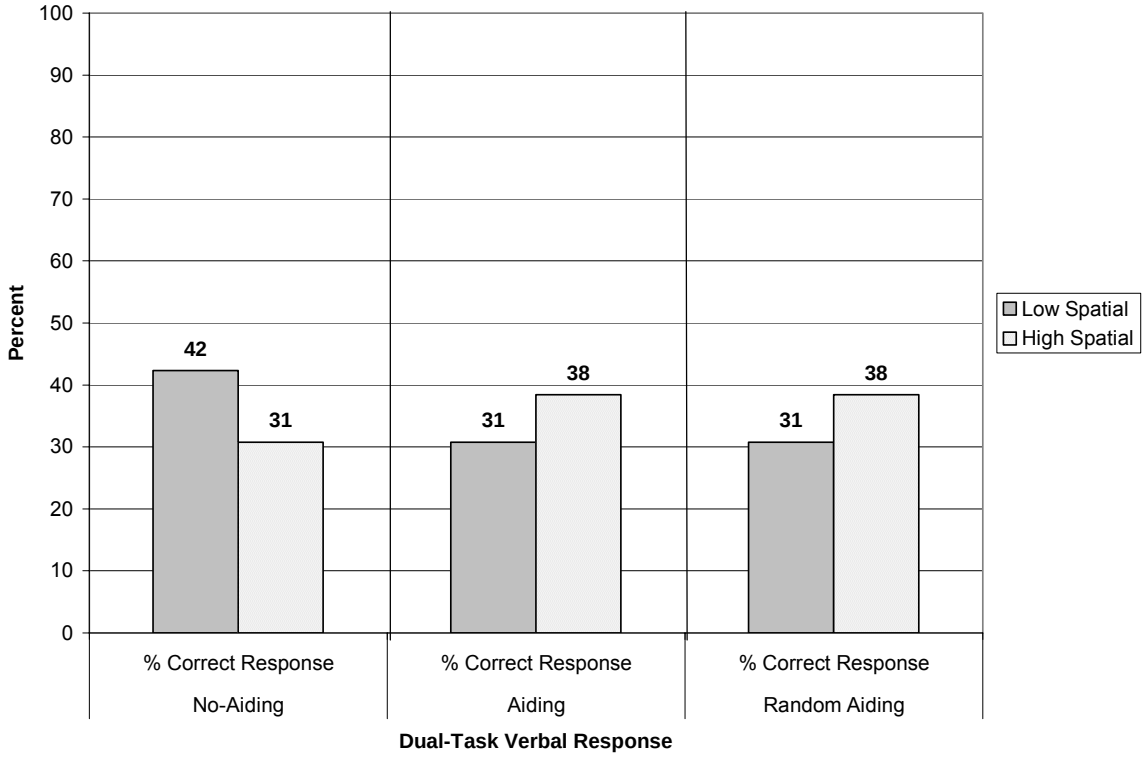

Figure 39. Percent correct responses for dual-task VHT performance analysis.

The effect of workload was as expected for the no-aiding trials. During the high workload trials, the operators had a 9% reduction in performance relative to the low workload trials. The effect reversed for the aided trials and the randomly aided trials.

The operators were briefed and trained that both the OVI task and VHT were equally important during the dual-task experiment. The VHT performance results could be caused by several conditions. The operators could have shed the VHT as the workload increased, resulting in the decreased performance. If the operators shed the task, the number of no responses should be high. Figure 40 shows the breakdown of incorrect

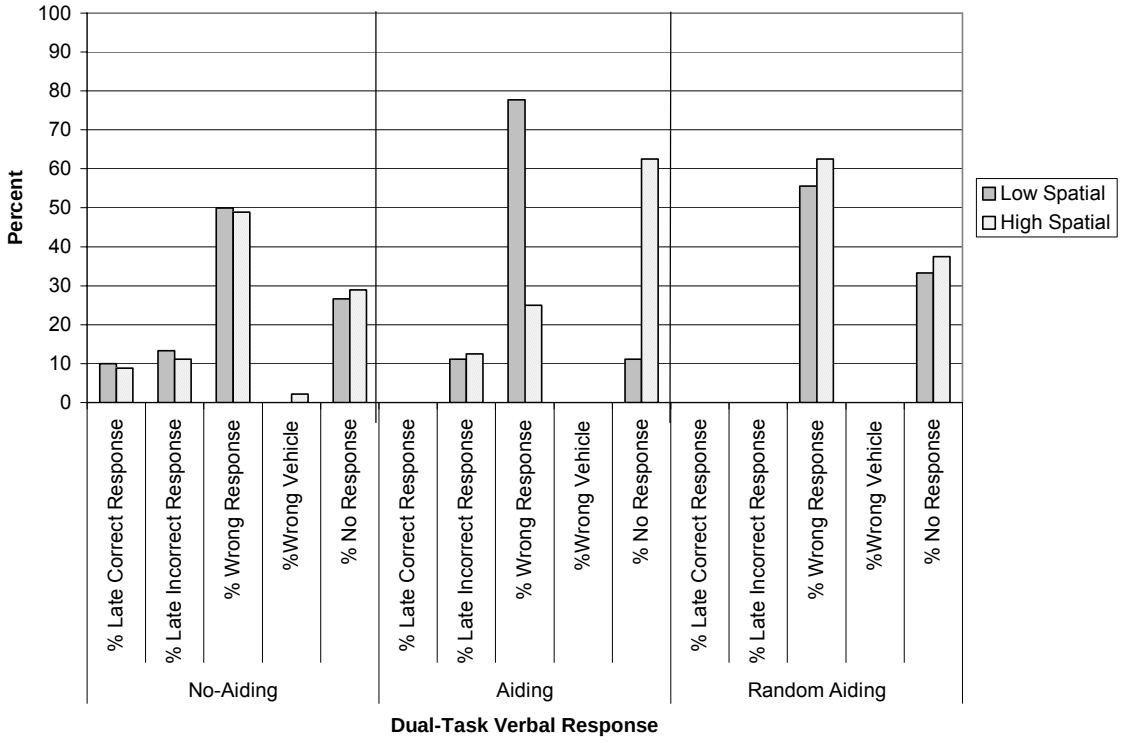

Figure 40. Breakdown of incorrect responses shows the majority of the missed responses are wrong responses.

operator responses. The majority of the incorrect responses are wrong responses. The results indicate that the operator did not shed the task but could not perform the VHT while maintaining proficient performance in the OVI task. Additionally, the aiding did not improve performance on the VHT, indicating that aiding must be specific to the task.

4.2.3 Online Classification

The online real time classification accuracy was 69.5% for both the low and high workload conditions. This accuracy is above chance but still not as high as expected. Even with low accuracy, the classifier triggered enough of the time to result in an increase in operator performance as discussed in Section 4.2.2.1. The training trials were randomly separated into training data, validation data, and test data. The test data was classified correctly in about 96% of the samples. The ANN did not overlearn the training data because the algorithm used the validation data to determine the optimum weights for generalization to the test data set.

The classification of the high workload condition was not as accurate (more misclassification) as the low workload condition. Since the classification accuracy for both the low and high workload conditions were not as high as expected, further investigation into the classifier outputs and the psychophysiological measures was conducted. Screen captures (Figures 41 and 42) were made of the state classifier at the end of a classification performance trial. The classifier switched from low to high to low during the high workload portions of the trials.

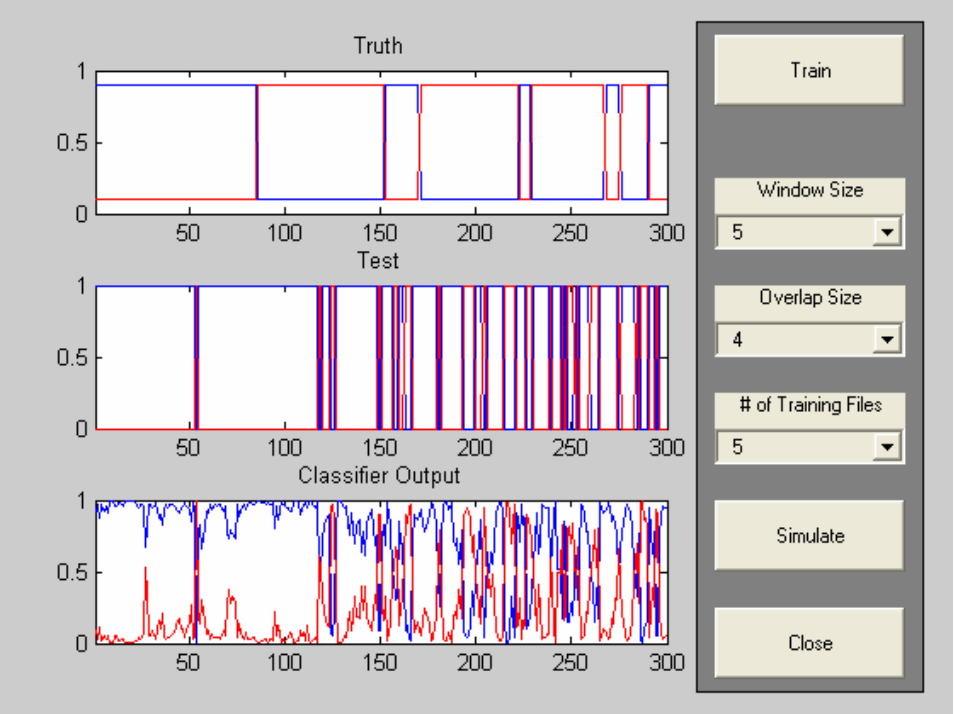

Figure 41. Screen capture of classifier results of a sample high workload trial. The test traces are the inputs to the system to determine aiding.

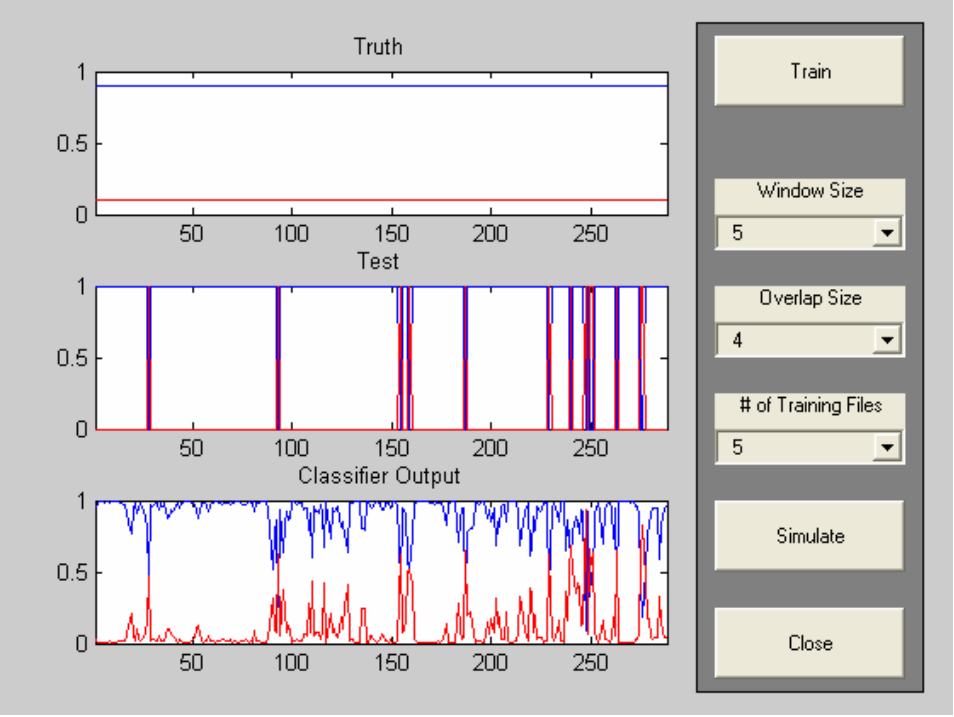

Figure 42. Screen capture of classifier results of a sample low workload trial. The test traces are the inputs to the system to determine aiding.

The classifier has two outputs, one each for the low and high workload conditions. The red trace represents high workload and the blue trace represents low workload. Three plots describe the figures. The top plot in each of the figures is the truth. The red trace is high for four periods. These periods are the times the SAR image is open and the operator is placing DMPIs on targets and represents high workload. The blue trace is high during the cruise or ingress to target and time between SAR images, which represents the low workload condition.

The center plot is the input to the system that determined the time when the system is aiding the operator. When the red trace is high, the system determines which vehicle the operator is currently attending and slows that vehicle down to allow the operator to complete the current task. When the blue trace is high, the system reverts to its previous state and increases the airspeed of the vehicle to the mission profile set during mission planning. The bottom plot is the actual output of the classifier output layer.

Figure 41 (high condition) shows that the classifier is switching back and forth during the high SAR conditions, not the baseline cruise condition. This oscillation may be due to the stability of input measures derived from the psychophysiological signals.

Figure 42 is a screen capture of the results of a low workload run (cruise and low SAR image). There are a few false alarms, but accuracy is still high. These results are from a typical subject. Several techniques could be used to improve the classification.

A 5-second window and a 4-second overlap were used to smooth the data before application of the ANN, which leads to the question of whether to pre- or post-smooth the data. That is, to increase the window size and smooth the data before training the

classifier, or use shorter windows and smooth the output of the classifier, or both. Both methods introduce delay in the overall response of the system.

Another consideration is that the workload is consistently high during the high SAR image processing intervals. Figure 43 is a plot of the T5 gamma magnitude and the truth for a high trial. Feature T5 gamma is a highly salient feature - one that is weighted higher in the operator function state model as determined by the weight based partial derivative saliency technique described in Section 2.7 (saliency is discussed in Section 4.6). The magnitude of the EEG gamma signal increases as the operator progresses further into the mission. In fact, the significant increase occurs about half way through processing the second SAR image, possibly due to the operator delaying processing

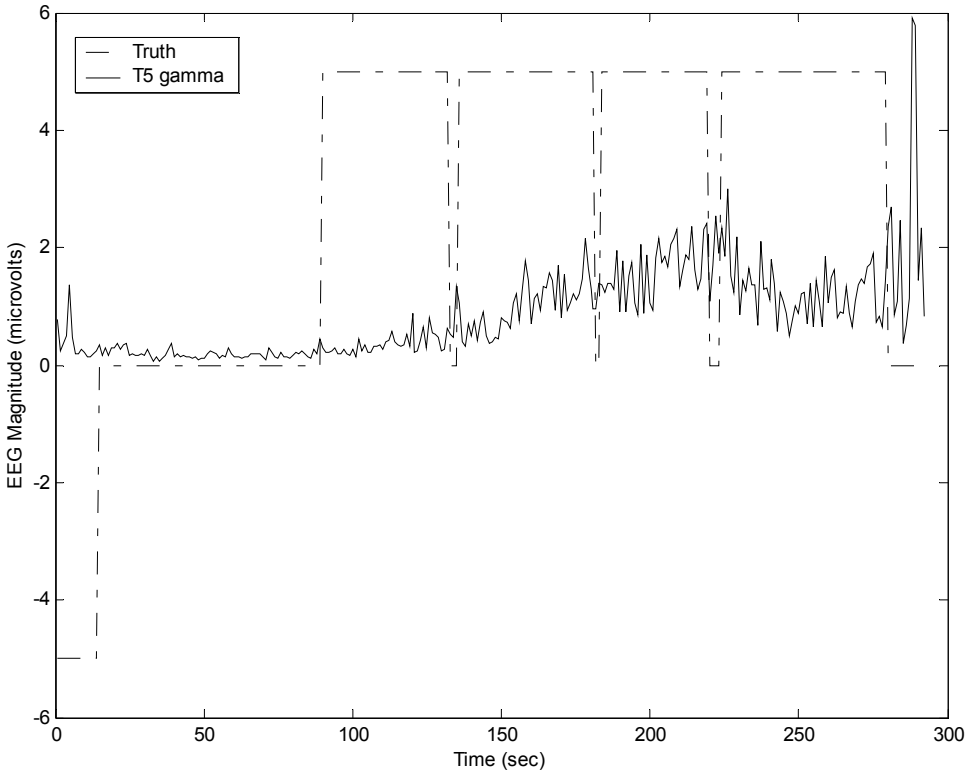

Figure 43. Output of T5 gamma during a high trial shows the magnitude increases well into the second SAR processing interval. The dashed line is high during the period the SAR image is open.

earlier SAR images and having to catch up upon realizing the remaining SAR images cannot be completed in time.

Additionally, operators have different skills and some can perform the high workload task with the same ease as in the low workload trials. In fact, one of the operators did not miss any targets regardless of aiding. The performance of the classifier for this operator was not as high as for the other subjects during the high workload periods of the trial. In fact, only 13% of the high workload period was classified as high workload. Operator physiology is not expected to change without cognitive loading. However, this concern is not a training issue as all operators were trained to the same performance level in the single-task experiment.

4.4 Classifier Comparisons

Several classifiers were compared using the data from the single-task experiment. Classification accuracy for each classifier was compared to classification accuracy of the ANN - the baseline algorithm for this study. Discriminant analysis and support vector machines were compared to the ANN using the same training and test data. Data sets were prepared as described in Section 3.5.

Classification accuracy using linear discriminant analysis, quadratic discriminant analysis, and logistic discriminant analysis techniques were compared to the results found using a multilayer perceptron with backpropagation training (Figure 44). Classification accuracies across all algorithms were similar, but the ANN tended to outperform the others.

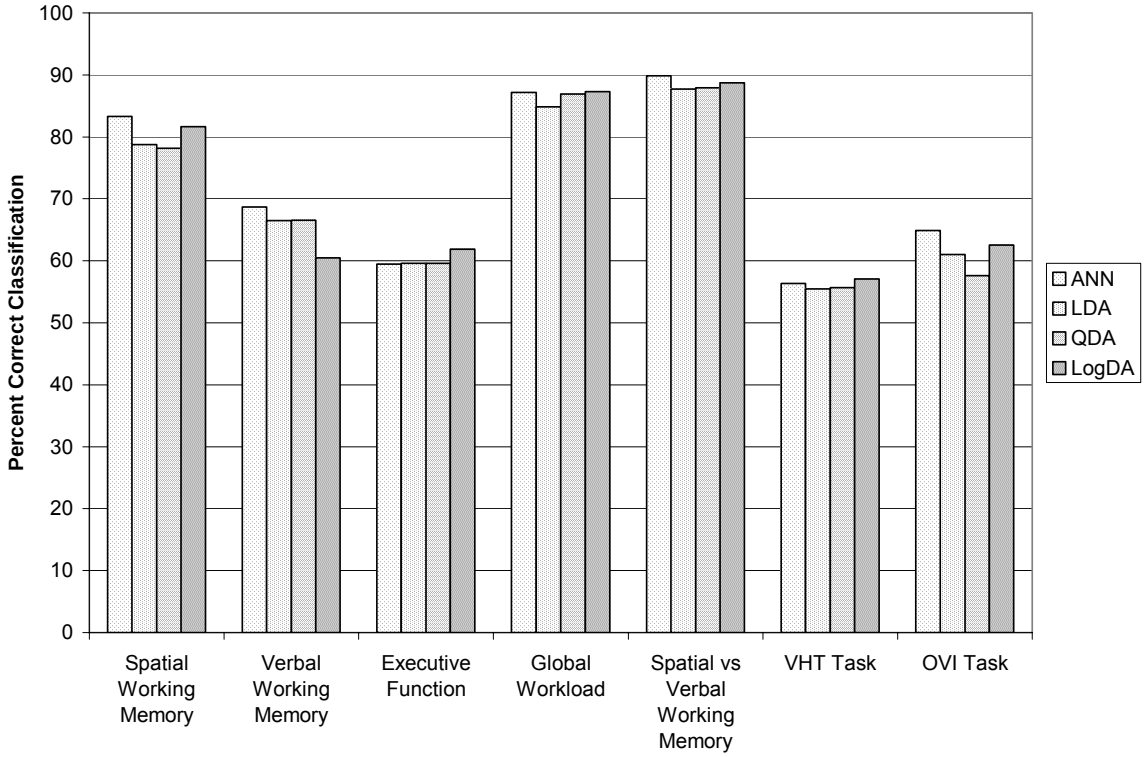

Figure 44. Classification accuracy for the artificial neural network was better as compared to discriminant techniques for most cognitive gauges.

Comparisons between the discriminant analysis techniques and the artificial neural networks were conducted pairwise since the artificial neural network was used as the baseline for this research. The wins for the ANN were summed and divided by the total number of trials. The wins were collapsed across cognitive gauge classification. The ANN performed better in each case. Figure 45 shows the ANN win percentages against each of the discriminant analysis classifiers. The worst performer was linear discriminant analysis (LDA), which lost to the ANN in 80% of the models. Quadratic discriminant analysis (QDA) did better with the neural networks winning 68% of the trials. The best performer against the ANN was logistic discriminant analysis (LogDA), which only lost 58% of the time.

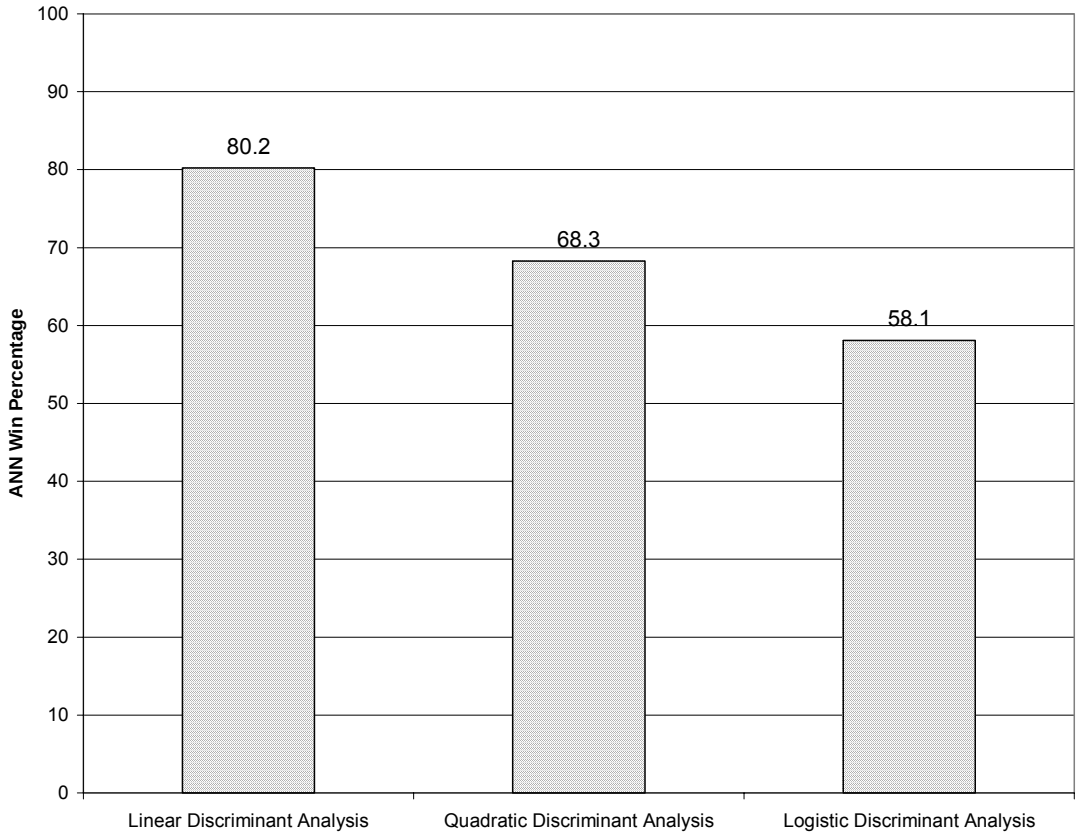

Figure 45. Win percentage of the artificial neural network classification over discriminant analysis techniques across all trials and cognitive gauges.

Comparisons between the discriminant analysis techniques and the artificial neural networks were conducted pairwise using McNemar's test as described in Section 2.12. McNemar's test can be compared to a chi-squared distribution with one degree of freedom as a test for the improvement in correct classification in classifier A versus classifier B. The ANN win pooled probabilities of these tests are shown in Figure 46. The win probability varied by cognitive gauge, but in each case the results have the same trend. The worst performer is LDA, followed by QDA discriminant analysis, then finally LogDA. Note the ANN wins by a lower margin as the discriminant analysis models become more nonlinear.

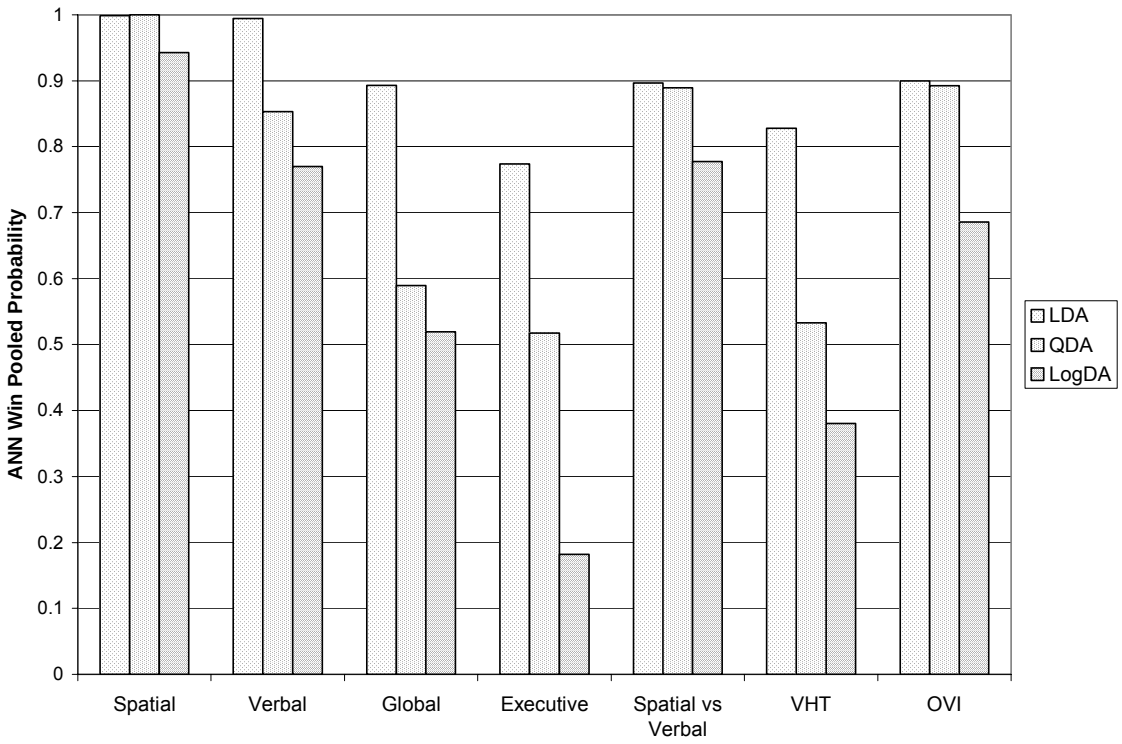

Figure 46. Artificial neural network pooled win probability for each of the cognitive gauges and discriminant classifier comparisons.

Comparisons between support vector machines (SVM) and artificial neural networks were also accomplished. Three support vector machines were evaluated - linear support vectors, polynomial support vectors, and radial basis function support vector machines. The linear support vector machine is a special case of the polynomial support vector machine. The kernel function for the polynomial learning machine is $\left(x^T x_i + 1\right)^p$, where p is specified by the user *a priori.* Figure 47 summarizes the classification accuracy for all cognitive gauges using polynomial support vector machines with orders of 1, 2, 3, 4, 5, and 6.

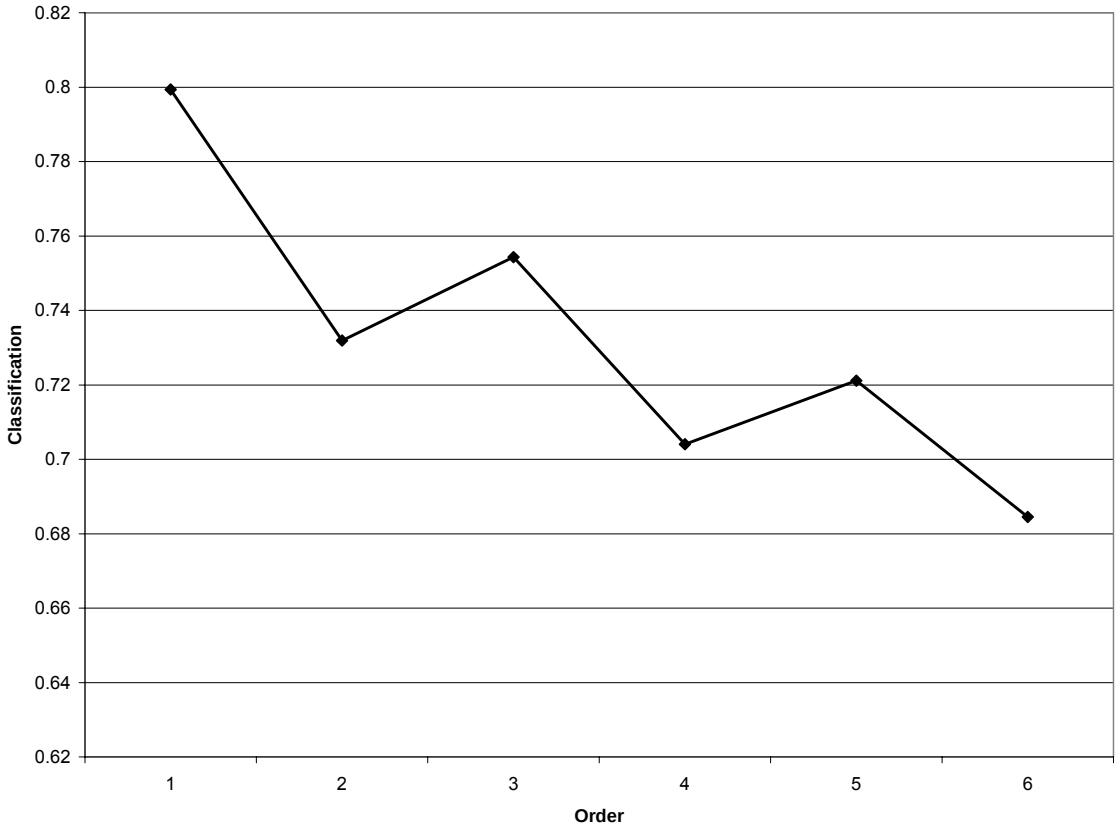

Figure 47. Polynomial order must be determined for the kernel in the polynomial support vector machine.

The best classification is with a polynomial of order one; however, the linear support vector machine was already in consideration. The next best order for the polynomial kernel is 3rd order. The kernel for the radial-basis function network is $\exp\left(-\frac{1}{2\sigma^2}\|x-x_i\|^2\right)$, where σ is specified by the user *a priori.* The spread of the radial basis function was determined in the same manner as the order was determined for the polynomial kernel. Figure 48 is a plot of classification accuracy using σ of 0.01, 0.05, 0.1, 0.25, 0.5 and 1.0. The best spread for the radial basis function was 0.05, and this value was used for the evaluation of the radial basis function SVM.

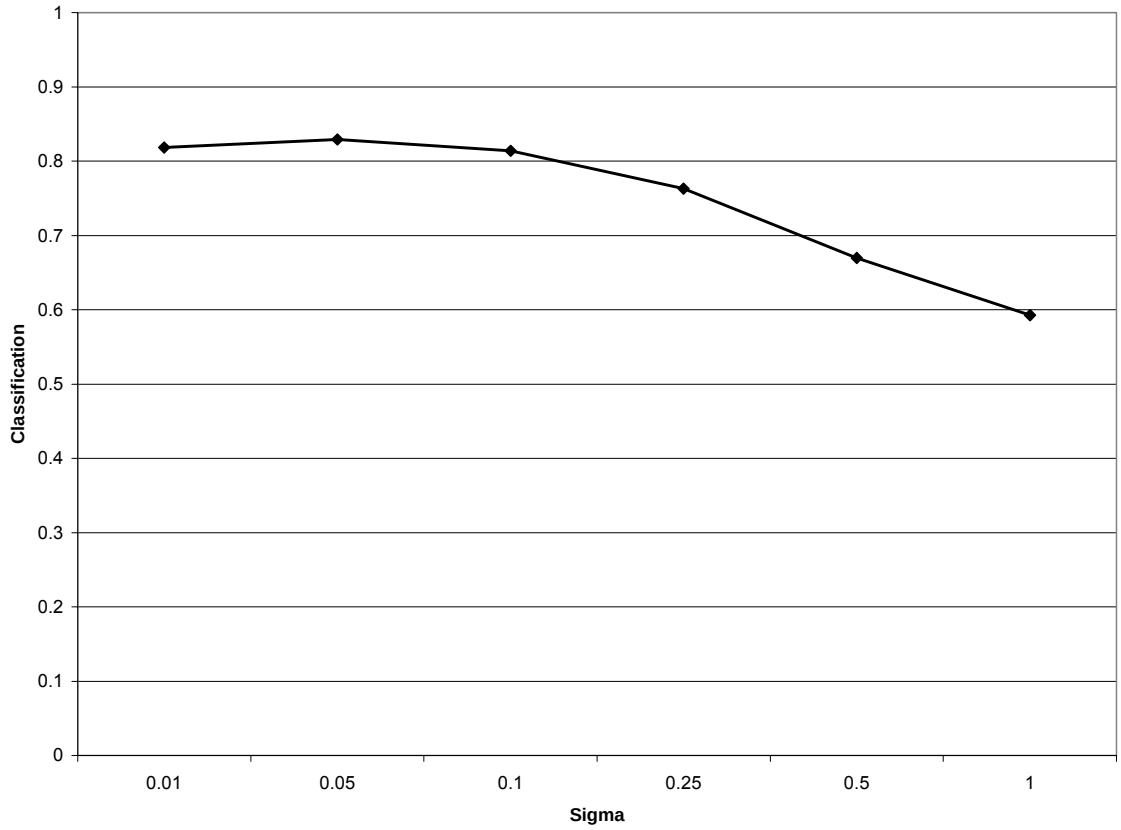

Figure 48. Radial basis function width must be determined for the kernel in the radial basis function support vector machine.

The results using SVMs were compared to the results obtained using the ANN. The classification accuracy for each algorithm is shown in Figure 49 for each cognitive gauge. As with the results using the discriminant functions, support vector machines have comparable classification accuracy to the artificial neural networks. However, support vector machines, particularly those using linear and radial basis function kernels, perform almost as well as the ANN. The 3rd order polynomial support vector machine did not perform as well as the other support vector machines. This result was expected since a 1st order polynomial was considered a better choice as a polynomial model based on the analysis to determine the order of the polynomial support vector machine.

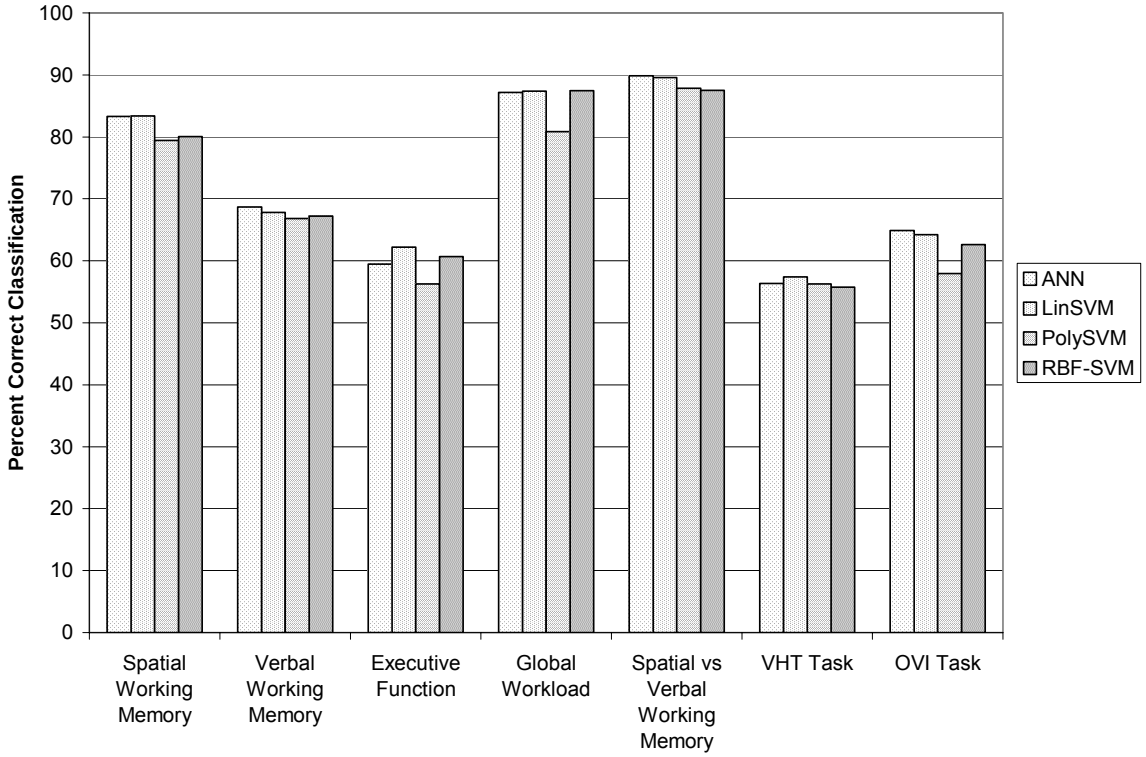

Figure 49. Classification accuracy for comparing results using support vector machines and artificial neural networks for each of the cognitive gauges.

Comparisons between the support vector machines and the artificial neural networks were conducted pairwise. The wins for the ANN were summed and divided by the total number of trials. The wins were collapsed across cognitive gauge classification. The ANN performed better in each case. Figure 50 shows the ANN win percentages against each of the support vector machine classifiers. The worst performer was 3rd order polynomial SVM, which lost to the ANN in approximately 76% of the trials. The linear SVM and the radial basis function SVM performed about the same, with the artificial neural networks outperforming these algorithms in about 59% of the trials. These results show that the ANN is the better algorithm for classifying operator functional state using psychophysiological measures.

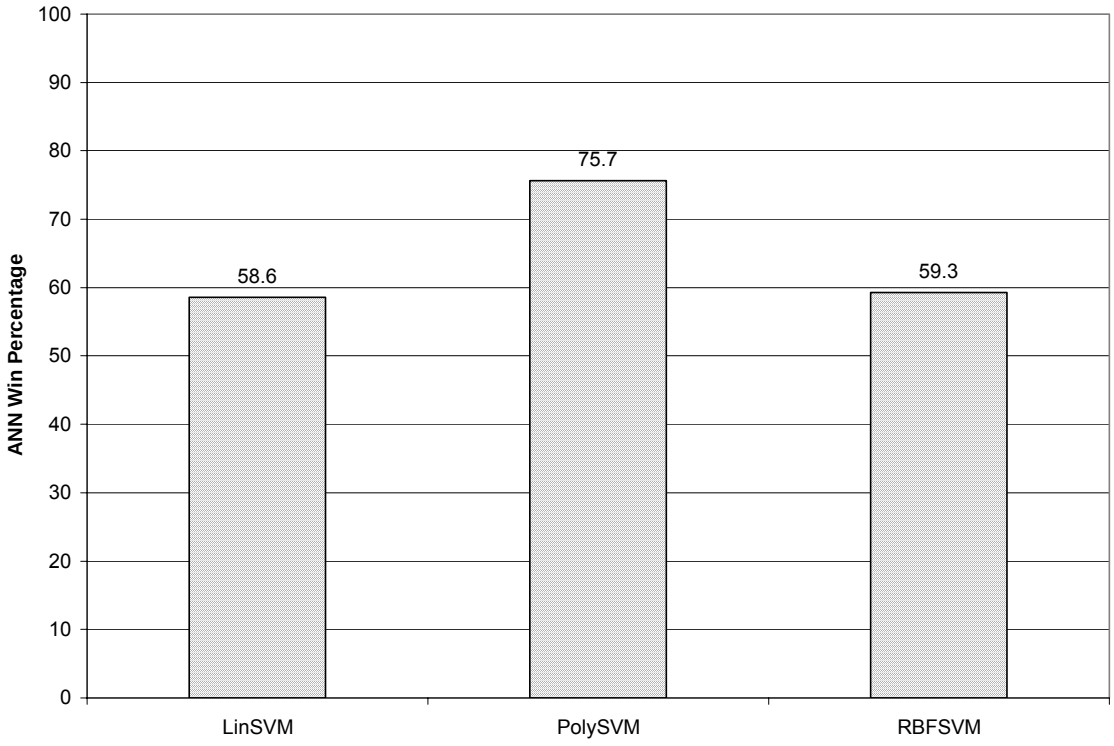

Figure 50. Win percentage of the artificial neural network classification over support vector machines across all trials and cognitive gauges.

Comparisons between the support vector machines and the artificial neural networks were conducted pairwise using McNemar's test described in Section 2.12. The ANN win pooled probabilities of these tests are shown in Figure 51. The win probability varied by cognitive gauge but did not follow the same trend as was the case for evaluation with discriminant functions. The worst performer was the 3rd order polynomial support vector machine. In fact, the artificial neural network outperformed the algorithm in every trial during the spatial working memory, global workload, and OVI task cognitive gauge trials. The ANN won by a lower margin with the linear learning machines and the radial basis function support vector machines.

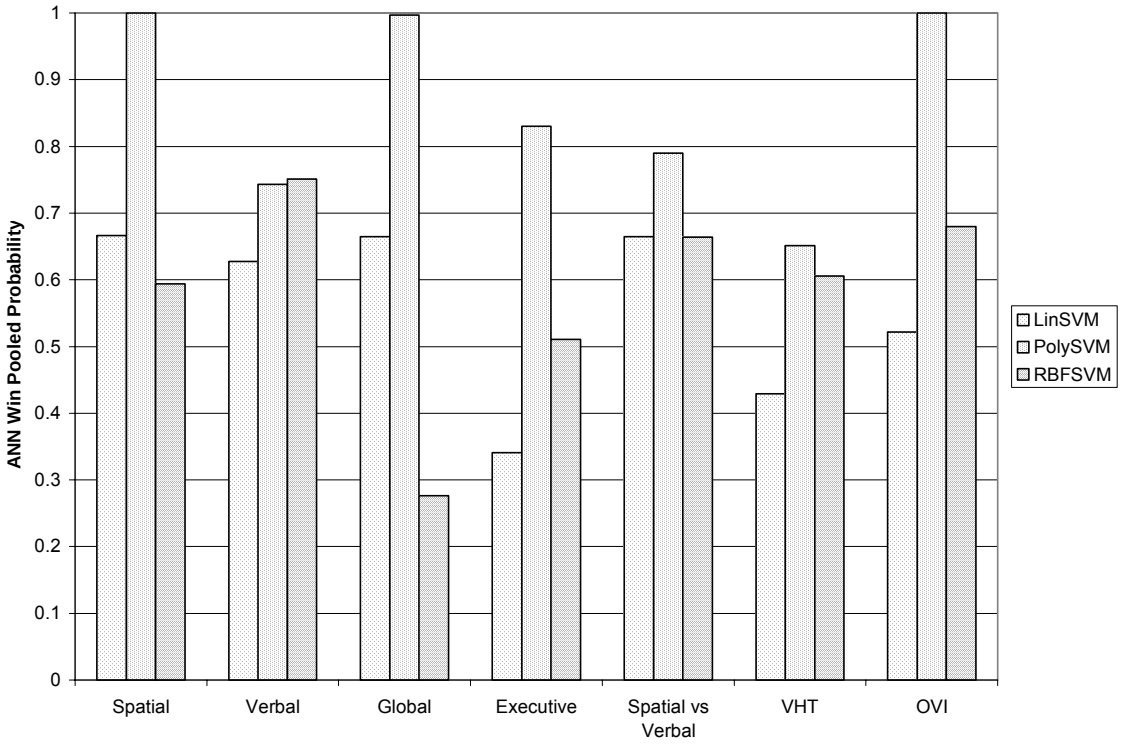

Figure 51. Artificial neural network pooled win probability for each of the cognitive gauges and support vector machine classifier comparisons.

Support vector machines have been demonstrated in many 'toy' data set problems that are Gaussian with large margin class boundaries. The support vector machines perform well and even better than the multilayer perceptron due to the underlying statistics of the data. However, the multilayer perceptron outperforms the SVM in most real-world problems. In fact, one researcher (Raudys, 2000) showed an increase in algorithm overlearning of the training sets and increases in margin width for determining the optimal hyperplane in many real world data sets, and also claimed that a specifically trained perceptron that is optimally stopped using validation data is a better alternative to a linear support vector machine.

4.5 Inclusion of New Measures-Classification Results

Neural networks were trained using the input features from the study (EEG, A-meter, heart, and eyeblink measures) as well as new measures collected offline during the trials. Electrodermal activity (EDA), electromyography (EMG), and pupil diameter were collected as described in Section 3.2. The EDA, EMG, and pupil data collected during the single-task experiment were used in addition to the features used to train the ANNs to determine improvements in classification, if any, by using the new measures collected offline. The data were trained using a multilayer perceptron with backpropagation with the same training and test data sets used in the single-task experiment. Figure 52 shows the results for this experiment for each of the cognitive gauges. The results are compared to the results obtained using the original features of the single trial experiment.

The additional measures improve overall classification accuracy. However, the increase is small, ranging from 0 to 6% with an average of about 2%. The additional cost of the equipment and the burden on the operators of adding the new equipment, electrodes, and more weight may outweigh the improvement in classification accuracy. Saliency analysis, discussed in the next subsection, can determine how important the measures are to accurate classification.

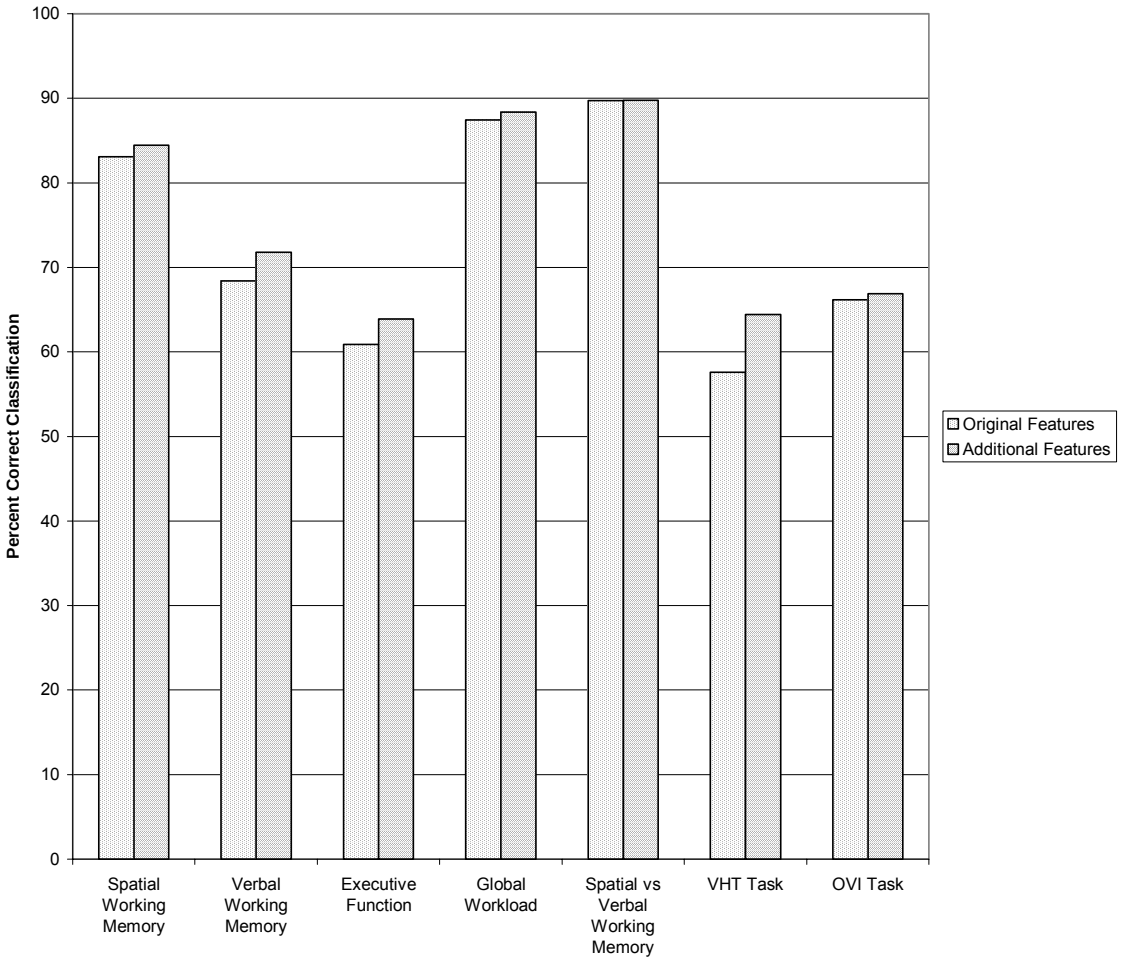

Figure 52. Classification accuracy improves with additional features, indicating that the features are salient.

4.6 Saliency Analysis

Saliency analysis was conducted using the partial derivative saliency measure and the trained neural networks from the experiment described in Section 4.5, which included the measures collected offline as well as those used in the real-time classification. The partial derivative technique computes an input-output relationship for each of the features using partial derivatives of the layer outputs in a fully trained network.

The saliency values for each of the trained networks was normalized so the most salient feature had a saliency of one and all other features had values less than one but maintained their relationships with all other features. The saliency was summed for each of the features for all trials for a particular cognitive gauge. The values were normalized in the same manner as the individual trials, which was accomplished for each of the cognitive gauges, and results can be found in Appendix D. The normalization procedure was accomplished to ensure that the features for each operator were in the same range of values to allow the saliency to be collapsed across trial and operator.

The saliency values are ranked in descending order by cognitive gauge, and the feature labels themselves are displayed in rank order with the most salient feature at the top of the table (see Appendix E). The top ten salient features for each of the cognitive gauges are displayed in Table 22. Many of the features are salient for each of the cognitive gauges, indicating that different measures are not necessary for different cognitive gauge classifiers.

The additional measures described in Section 4.5 appear as salient features in all of the cognitive gauges. In particular, the pupil diameter ranks in the top four for all seven cognitive gauges, indicating it is important to classification. The majority of the EEG features are the higher frequency measures of beta and gamma. They are also associated with electrodes around the edge of the scalp and so may be measures of tonic muscle activity.

Table 22: Top ten salient features for each of the cognitive gauges for all operators.

Spatial Working Memory	Verbal Working Memory	Executive Function	Global Workload	Spatial vs Verbal	VHT Task	OVI Task
Pupil Diam	EDA Tonic	EMG	HEOG gamma	Interblink	EDA Tonic	Pupil Diam
Interblink	EMG	EDA Tonic	Pupil Diam	EMG	EMG	EMG
T5 gamma	Pupil Diam	Pupil Diam	T5 gamma	O2 alpha	Pupil Diam	EDA Tonic
O2 alpha	O2 gamma	T5 gamma	EDA Tonic	Pupil Diam	O2 gamma	Interblink
F7 gamma	F7 gamma	O2 gamma	O2 gamma	O2 gamma	F7 gamma	HEOG gamma
EMG	T5 gamma	HEOG gamma	EMG	T5 gamma	T5 gamma	Interbeat
EDA Tonic	T5 beta	F7 gamma	F7 gamma	EDA Tonic	VEOG gamma	T5 beta
O2 gamma	FZ gamma	T5 beta	Interbeat	VEOG gamma	F7 beta	F7 gamma
VEOG gamma	Interblink	Interblink	Interblink	HEOG gamma	T5 beta	O2 gamma
HEOG gamma	VEOG gamma	VEOG gamma	VEOG gamma	O2 theta	Ameter	T5 gamma

V. Discussion and Conclusions

5.1 Overview

This chapter discusses the results of each experiment and considers conclusions that follow from them. The single-task experiments were conducted to evaluate and compare classification algorithms for operator functional state. These experiments were also used to evaluate the ability to develop cognitive models derived from information processing demands and task type. Additionally, these experiments were used to evaluate new, nontraditional psychophysiological measures and their utility in improving classification accuracy. The dual-task experiments were conducted to determine the utility of adaptive aiding using operator functional state in a UCAV simulation.

5.2 Single-Task Experiment Discussion and Conclusions

Single-task experiments were conducted to explore three questions concerning operator functional state estimation: 1) Can multiple cognitive gauges be developed based on information processing demands and task type? 2) Which pattern classification algorithm works best for classifying operator functional state using psychophysiological measures? 3) Which psychophysiological measures are salient in classifying operator functional state? Each question is considered in the next three subsections.

5.2.1 Multiple Cognitive Gauge Development

Multiple cognitive models or gauges were developed based on information processing demands and task type. Some models were more accurate than others using the psychophysiological measures investigated in these experiments. Models developed based on information processing demand were for global workload, executive function,

spatial working memory, verbal working memory, and spatial versus verbal working memory. Additional models were developed for the OVI and VHT tasks.

A single-task experiment was conducted to explore implementation of different cognitive gauges using the same psychophysiological measures. Seven cognitive gauges: verbal working memory, spatial working memory, global workload, executive function, spatial versus verbal working memory, and an OVI task and a VHT gauge were evaluated. Artificial neural networks, specifically multilayer perceptrons, were used to train each gauge. Results showed that different gauges could be determined using the same features and that classifications for some gauges have much better performance. The gauge to determine spatial or verbal working memory performed best with a classification accuracy of about 91%. The accurate classification of this gauge indicates that features derived from physiological signals can be used to differentiate classes of cognitive processing accurately, as in the case of working memory.

The verbal working memory and VHT gauges, however, did not perform well. The separation between the low VHT (low verbal working memory) and the high VHT (high verbal working memory) classes was not sufficient to distinguish between the two classes. Operator subjective measures calculated using the NASA-TLX showed significant differences between these two classes. Operator performance also showed significant differences between the classes. The classifiers were expected to find differences in the physiology, since differences were found in both subjective measures and operator performance measures. The poor separation could be a result of the current location of the EEG electrodes, which may not record signals from regions of the brain

responsible for verbal working memory. Future studies should consider these regions when determining sensor placement.

The executive function gauge also had low classification accuracy (70%), although the accuracy was well above chance (33%). The trial segments defining the levels of executive function were derived from the mental demand subscale of the NASA-TLX. The low classification accuracy could result from an inadequacy of mental demand subscale as a good measure of executive function and the location of the EEG sensors. However, the latter is unlikely since executive function occurs in the frontal lobe of the brain, and two sensors in this study were located in the frontal region (F7 and Fz).

The spatial working memory and OVI tasks were based on the SAR image processing and ingress portions of the study. The classifiers performed fairly well (81% for spatial working memory and 70% for the OVI task) for each of these gauges. The operator subjective ratings and operator performance measures also showed significant differences in the low and high conditions. The gauge for the global workload for real-time classification in the dual task experiment was derived from these two single-task cognitive gauges.

5.2.2 Pattern Classification Algorithm Comparison

The data from these experiments were also used to explore the utility of various pattern classification algorithms. An evaluation of classifier algorithms was conducted to determine which classifiers performed better using psychophysiological signals in a complex operational environment. Comparisons were made between artificial neural networks, discriminant functions, and support vector machines. The artificial neural networks performed better; however, other algorithms could be considered adequate

substitutes. Support vector machines performed well with psychophysiological signals, particularly in the case of linear and radial basis function support vector machines.

Other issues should be considered when selecting the classifier. The artificial neural network always trained to the data with near perfect results, but it could overlearn the data and not generalize to new data samples. If the developer does not realize this effect and does not implement techniques such as early stopping using validation data, the artificial neural network may not perform as well as other algorithms that are properly trained. Similar issues apply to any algorithm.

Algorithm complexity should be a consideration; Occam's razor selects the simplest solution as the best solution. More complex algorithms have more parameters that must be determined. Also, an increase in the number of parameters means that more training data must be collected to build a good model of cognitive workload.

5.2.3 Psychophysiological Feature Saliency

Feature saliency was explored using this data set. The partial derivative saliency method was used to determine the relative importance of features in model accuracy. New measures, such as integrated muscle activity, arousal level, pupil diameter, and electrodermal tonic level, were evaluated in addition to traditional psychophysiological features. Feature saliency analysis indicated that the same features can be used to detect levels in multiple cognitive gauges. The single-task experiment showed that the same features appeared in each of the cognitive gauge top ten list. Feature saliency can help prune the input features used for classification. Reducing the number of features decreases algorithm training time and reduces the complexity of the classifier. Reducing the number of signals collected increases operator acceptance of this new technology.

5.3 Dual-Task Experiment Discussion and Conclusions

Adaptive automation using operator functional state was demonstrated to improve mission effectiveness by decreasing the number of missed weapons release waypoints and increasing the number of targets hit. These experiments represent the first implementation of this technology in an operationally relevant environment.

5.3.1 Utility of Operator Functional State

A dual-task experiment was conducted to determine the utility of operator functional state derived from psychophysiological signals in adaptive aiding to improve operator performance. Operator performance was evaluated based on signal detection measures derived from SAR processing and mission effectiveness measures. Several trials were conducted to evaluate the ability of artificial neural networks to detect a high workload condition in the operator. The classifier did not perform as well as expected; the operator state was classified correctly with 70% accuracy. The low classification accuracy may have several causes, the most obvious of which are the input features. The input measures may not be robust enough to classify operator state, but the lack of robustness is probably not the cause since studies conducted by numerous researchers have shown promise for classifying operator functional state as described in Sections 2.4.4 and 2.4.5.

Another consideration is consistently high workload during the high workload intervals and low workload during the low workload intervals. The results in Section 4.2.3 (describing the dual-task experiment) showed that the magnitudes of one of the most salient features were not consistent during the labeled high workload condition. The magnitudes of the EEG gamma signal increased as the operator progressed further into

the mission, which could indicate that the operator workload was not consistent during the high workload task segments. Operator performance during the high workload trials seems to provide further evidence that the workload levels were not consistent during the processing of the SAR images. All operators were able to complete the task of processing the first two SAR images before the weapons release points. However, each operator did not complete at least one of the third and fourth SAR images, resulting in a missed weapons release and a mission failure for the associated UCAV.

Additionally, operators had different skill levels, and some can perform the high workload tasks with the same ease as the low workload trials. In fact, one of the operators did not miss any targets regardless of aiding. Operator physiology is not expected to change without cognitive loading. In fact, the classifier for this operator estimated only 13% of the high workload condition in the trial as high workload. This result is not a training issue, as all operators were trained to the same performance level in the single task experiment.

Since operators perform at different levels and their cognitive load increases as a result of increased task demands, future studies should include experiments that define the operator workload based on performance. These studies could include trials of increasing task demand. The high workload trials could consist of simulation parameters for each operator based on the point of individual performance breakdown. The trials in the experiment could be tailored for each operator.

5.3.2 Manipulations of Operator Vehicle Interface

The dual-task experiment also investigated effects on operator performance using operator functional state for adaptive aiding by manipulating the operator vehicle

interface task. The adaptive aiding consisted of slowing down the vehicle that the operator was focused on when the classifier detected the operator was in high cognitive workload. The aided trials showed a significant increase in target hits and consequently fewer missed targets. The mission effectiveness parameters also showed significant improvement in the adaptively aided trials. The number of vehicles that missed the weapons release waypoint was reduced by 67% by aiding the operator adaptively. This result occurred in spite of the classifier being correct only 70% of the time. Improved classification accuracy could further improve mission effectiveness. These results are critical data in operational settings since missed weapons release points are wasted missions.

5.3.3 Time-appropriate and Task-appropriate Aiding

Some trials were implemented using a randomly aided scheme in which the operator was aided regardless of functional state. The results indicated that the aiding must be presented at the appropriate time. The operator performance for mission critical measures (number of target hits and number of missed weapons release waypoints) during the randomly aided trials was not significantly different from the trials with no aiding. In addition to presenting adaptive aiding at the appropriate time, the results suggest that the aiding must be appropriate for the task. No significant differences in vehicle health task performance were found in no-aiding trials, aided trials, or in the randomly aided trials.

VI. Summary and Recommendations

6.1 Overview

This dissertation makes contributions toward increasing operator performance using adaptive automation and operator functional state. This chapter re-emphasizes the significant contributions of the research. Recommendations for continuing research are also presented.

6.2 Significant Contributions

Several firsts were presented in this dissertation. This research was the first example of adaptive aiding using operator functional state in an operationally relevant environment. Measures derived from psychophysiological signals were identified, extracted, and integrated using the multilayer perceptron. The output of the multilayer perceptron defined the operator functional state and was used as a control input to the UCAV simulator. In turn, the simulator enabled mitigation strategies during high cognitive workload periods, allowing the operator to focus on mission critical events. Once the operator functional state was determined by the pattern classifier to be at a nominal level, the system was returned its previous state.

Implementation of adaptive aiding using operator functional state resulted in improved operator performance. Improvements in mission critical performance measures were significant. For example, implementation of operator-functional-state-driven adaptive aiding reduced the occurrence of missed weapons release waypoints, a measure of mission failure, from 25% in the trials without adaptive aiding to 8% in trials with adaptive aiding. This result represented a performance improvement of 67% ± 3%.

Another first presented in this research was the development of multiple cognitive models using the same psychophysiological measures. The multiple cognitive models were defined by information processing demands and task type. Previous experiments focused on one information processing demand, i.e, working memory or global workload. This research developed seven cognitive models, five defined by information processing demand and two derived from the simulation task.

Finally, multiple pattern classification algorithms were compared to determine their utility for classifying operator functional state using psychophysiological measures. Three types of pattern classification algorithms: support vector machines, discriminant analysis, and artificial neural networks were explored. The multilayer perceptron neural network classifier was found to be marginally superior.

This research has resulted in several publications and presentations. Four papers were published in journals and conference proceedings. Portions of this research appeared in a NATO technical report.

6.3 Recommendations for Future Research

Future research should focus on improving the classification of operator functional state. Improvements in the classifier should improve operator performance since the assessment of operator state will be more accurate. The classifier algorithms used in this study all performed well, but classification accuracy was a limiting factor in operator performance improvement. Developing new input features or applying different transformations to existing measures could improve classification accuracy. Some transformations to explore are coherence between the features and using relative power instead of log power.

Future research should also investigate techniques and measures to predict cognitive load, not merely to identify current state. Cognitively overloaded operators perform adequately for short periods by focusing more on a task, but they may miss important time critical information because of this focus and may not be completely aware of other events in a mission. Determining that an operator is approaching an overload condition could prevent the onset of performance and situation awareness degradation. Implementing adaptive aiding before errors occur should improve mission effectiveness.

The multilayer perceptron with backpropagation training is a memoryless classifier, whereby the previous state or states are unknown to the classifier. The use of recurrent neural networks may enable the use of temporal information as well as spatial information. Temporal and spatial information may be used to predict operator state. Examples of recurrent neural networks are Elman neural networks, Jordan networks, and time delay neural networks (Haykin, 1999). Future studies should investigate the use of these artificial neural networks using psychophysiological data for predicting operator functional state.

Appendix A. Vehicle Health Task Failure/Correct Response Pairings

CATEGORY	PROMPT	APPROPRIATE RESPONSE
Electrical		
	Battery Power Low	Switch to Back-up Batteries
	Generator Fault Detected	Recycle Generator
Mechanical		
	Bomb Bay Door Fault	Recycle Bomb Bay Door
	Weapon Release Actuator Fault	Recycle Weapons Release Actuator
Engine		
	Engine Temperature High	Open Air Cooling Intake
	Engine Fault Detected	Check Fault Code
Sensors		
	SAR System Fault	Re-intialize SAR System
	GPS Signal Failure	Re-intialize GPS System
Communications		
	Loss of Communications	Switch to Alternate Comm Frequency
	Loss of Last Transmission	Re-send Last Transmission
Fuel		
	Fuel Load Unbalanced	Rebalance Fuel Tanks
	Fuel Pump Fault	Switch To Reserve Fuel Pump

Appendix B. Vehicle Health Task Command and Response Matrix

Command

Electrical	Mechanical	Engine	Sensors	Communications	Fuel
Battery Power Low	**Bomb Bay Door Fault**	**Engine Temperature High**	**SAR System Fault**	**Loss of Communications**	**Fuel Load Unbalanced**
Generator Fault Detected	**Weapon Release Actuator Fault**	**Engine Fault Detected**	**GPS Signal Failure**	**Loss of Last Transmission**	**Fuel Pump Fault**

Response

Electrical	Mechanical	Engine	Sensors	Communications	Fuel
Re-route Power	**Recycle Bomb Bay Door**	Restart Turbine	**Re-intialize SAR System**	**Re-send Last Transmission**	Dump Fuel
Recycle Generator	Recycle Landing Gear	**Check Fault Code**	Test Passive Radar Detector	**Switch to Alternate Comm Frequency**	Test Fuel System Pressure
Test Circuit Continuity	Recycle Flap Actuator	**Open Air Cooling Intake**	**Re-intialize GPS System**	Test Carrier Frequency	**Switch To Reserve Fuel Pump**
Switch to Back-up Batteries	**Recycle Weapons Release Actuator**	Reduce RPM	Recalibrate Altimeter	Recycle AWACS Communications	**Rebalance Fuel Tanks**

Correct response are **bold**

Appendix C. Kruskal-Wallis Nonparametric Test

The nature of the operator performance data dictated the application of nonparametric statistics. In some of the present data, normality could not be assumed. For example, the key measure for mission effectiveness (missed weapons release waypoints) is a series of Bernoulli trials which has a binominal distribution. An underlying assumption of the familiar Analysis of Variance (ANOVA) test is that the data have a normal distribution. Absence of normality in the operator performance data implies the need for the application of a nonparametric test. A common nonparametric test in human research is the Kruskal-Wallis test (Rosner, 1995; Siegel, 1956).

The Kruskal-Wallis tests the null hypothesis that *k* groups come from the same population with respect to the means. This test is used in place of the traditional ANOVA when the distribution of the sample data is not normal or the data are ordinal. The procedures for comparing the means using nonparametric methods are fairly straightforward and are outlined in the following steps.

1) Pool the observations from all groups, constructing a data set with a combined sample size of $N = \sum n_i$ where *n* is the sample size of the i^{th} group.
2) Replace the each of the *N* observations with ranks. The smallest observation is replaced by rank 1, the next smallest by rank 2, and the largest value by rank N. In case of ties, use the average rank of the tied observations.
3) Compute the rank sums for each of the groups.
4) Compute the test statistic. If there are no ties in the rank sums, the test statistic is

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1), \tag{83}$$

where R_i is the rank sum of the i^{th} group. If there are tied rank sums, the test statistic is

$$H = \frac{\frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N+1)}{1 - \frac{\sum_{m=1}^{k'}\left(t_m^3 - t_m\right)}{N^3 - N}}, \tag{84}$$

where t_m is the number of observations with the same value in the m^{th} cluster of tied observations and k' is the number of tied groups.

5) The H statistic used in the Kruskal-Wallis test is distributed approximately as chi square with $df = k - 1$. Test the null hypothesis H_0 that the group means are the same using

$$\begin{aligned} H &> \chi^2_{df,1-\alpha} \quad \text{Reject } H_o \\ H &\le \chi^2_{df,1-\alpha} \quad \text{Accept } H_o. \end{aligned} \tag{85}$$

6) Determine statistical significance by computing the p value

$$p = \Pr(\chi^2_{df} > H). \tag{86}$$

The Kruskal-Wallis test computes if groups are significantly different. Further testing must be accomplished to determine which groups are significant. Pairwise group comparisons can be made under the Kruskal-Wallis test to determine which group means are different using the Dunn Procedure (Rosner, 1995). To compare two groups, i_1 and i_2 use the following procedure:

1) Compute the z score, which is the test statistic for the Dunn Procedure, as

$$z = \frac{\overline{R}_{i_1} - \overline{R}_{i_2}}{\sqrt{\frac{N(N+1)}{12} * \left(\frac{1}{n_{i_1}} + \frac{1}{n_{i_2}}\right)}}, \tag{87}$$

where $\overline{R}_i$ is the average rank sum for group n_i. The Dunn Procedure adjusts the level of significance for the test for multiple comparisons. The level of significance or α is adjusted by

$$\alpha^* = \frac{\alpha}{m(m-1)}\ , \tag{88}$$

where m is the number of pairwise test being conducted and α is usually 0.05.

2) Use the z score and test the null hypothesis H_0 as

$$\begin{aligned} H &> |z| \quad \text{Reject } H_o \\ H &\leq |z| \quad \text{Accept } H_o \end{aligned} \tag{89}$$

3) Determine statistical significance by computing the p value

$$p = \Pr(|z| > H)\,. \tag{90}$$

This procedure is illustrated with an example using an operator performance measure from the research discussed in this document. The Missed Weapons Release Waypoint metric is a measure of mission effectiveness - a critical finding in this research. The Missed Weapons Release Waypoint measure is a success/failure metric and is scored as a 0 (mission success) or a 1 (mission failure). The complete raw data set for this measure can be found in Appendix G.

The first step of the Kruskal-Wallis test is to pool the raw data from all workload and aiding groups. The observations are sorted in ascending order and replaced with ranks. The observed values are 0 or 1 and the number of observations is 192. Obviously there are tied observed values in the data. The rank for observed values of 0 is $\frac{1}{170}\sum_{i=1}^{170} i = 85.5$ and for observed values of 1 the rank is $\frac{1}{22}\sum_{i=171}^{192} i = 181.5$. The ordered observed values and ranks are shown in Table 23.

Table 23: Table of Observed Value and Ranks for Missed Weapons Release Waypoint Measure.

Group	Observed Value	Rank
Low Workload No Aiding	0	85.5
Low Workload No Aiding	0	85.5
Low Workload No Aiding	0	85.5
. . .	. . .	. . .
High Workload Random Aiding	1	181.5
High Workload Random Aiding	1	181.5
High Workload Random Aiding	1	181.5

The next step in the procedures for the Kruskal-Wallis test is to compute the rank sums for each group. To improve the readability of the equations, the groups were assigned numbers. The low workload no-aiding group was assigned a 1, low workload aiding group 2, low workload random aiding group 3, high workload no-aiding group 4, high workload aiding group 5, and high workload random aiding was assigned to group 6. The rank sums for the groups are computed as

$$R_1 = 85.5 + 85.5 + \cdots + 181.5 + 85.5 = 4296. \tag{91}$$

Similarly, the ranks sums for the remaining groups are

$$R_2 = 2052, \tag{92}$$

$$R_3 = 2052, \tag{93}$$

$$R_4 = 5256, \tag{94}$$

$$R_5 = 2244, \tag{95}$$

and

$$R_6 = 2628. \tag{96}$$

Now the *H* statistic can be computed. Since there are ties, Equation (84) must be used to compute the *H* statistic as

$$H = \frac{\frac{12}{192*193}*\left(\frac{4296^2}{48}+\frac{2052^2}{24}+\frac{2052^2}{24}+\frac{5256^2}{48}+\frac{2244^2}{24}+\frac{2628^2}{24}\right)-3*193}{1-\frac{\left(170^3-170\right)+\left(22^3-22\right)}{192^3-192}} \tag{97}$$

$$H = 21.8578\,. \tag{98}$$

From a table of χ^2 values found in many statistics texts and using an α of 0.05 with $df = 5$, the critical χ^2 value is 11.070. The null hypothesis H_0 can be now be tested as

$$21.858 > 11.070 \quad \text{Reject } H_o \tag{99}$$

indicating the group means are significantly different with a probability of $p < 0.05$. Computing the p value directly from Equation (86) yields $p = 0.0006$.

To determine which groups are significantly different, pairwise comparisons of the groups using the Dunn Procedure for the Kruskal-Wallis test must be accomplished. First, compute the average rank sums. The average rank sum for group 1 (No-aiding under low workload condition) is

$$\overline{R}_1 = \frac{R_1}{n_1} = \frac{(85.5+85.5+\cdots+181.5+85.5)}{48} = 89.5\,. \tag{100}$$

Similarly, the average rank sums for the other groups are computed and the results are presented in Table 24.

Next, the test statistic *z* must be computed using Equation (87) for each pair of groups of interest. The first comparison is between no-aiding and aiding for the low workload conditions. The *z* score for this comparison is

Table 24: Table of Average Rank Sums for Missed Weapons Release Waypoint Measure.

Class	Average Rank Sum
Low Workload No Aiding	$\overline{R}_1 = 89.5$
Low Workload Aiding	$\overline{R}_2 = 85.5$
Low Workload Random Aiding	$\overline{R}_3 = 85.5$
High Workload No Aiding	$\overline{R}_4 = 109.5$
High Workload Aiding	$\overline{R}_5 = 93.5$
High Workload Random Aiding	$\overline{R}_6 = 109.5$

$$z_{12} = \frac{89.5 - 85.5}{\sqrt{\frac{192 * (192+1)}{12} * \left(\frac{1}{48} + \frac{1}{24}\right)}} = 1.1517 \tag{101}$$

Similarly, the other comparisons of interest are $z_{13} = 1.1517$, $z_{23} = 0.0$, $z_{45} = 4.6068$, $z_{46} = 0.0$, and $z_{56} = -4.6068$. The null hypothesis is tested using an α of 0.05 which is adjusted for multiple tests using (Equation 88) results in $\alpha^* = \frac{0.05}{6(5)} = 0.0017$. Finally, using α* and a table of z scores yields

$$\begin{aligned} H &> 3.14 \quad \text{Reject } H_o \\ H &\leq 3.14 \quad \text{Accept } H_o. \end{aligned} \tag{102}$$

Comparing the z scores for each of the pairwise comparisons indicates that only z_{45} and z_{56} are significantly different with a probability of $p < 0.05$. These groups are aiding versus no-aiding under the high workload condition and aiding versus random aiding under the high workload condition. Finally, the actual p values are computed using Equation (90) and those p values in addition to the z score are the values reported in Tables 18 and 19 in Section 4.2.2.1.

Appendix D. Saliency Values of Features for Each Cognitive Load Grouping

Feature	Spatial Working Memory	Verbal Working Memory	Executive Function	Global Workload	Spatial vs Verbal	VHT Task	OVI Task
HEOG delta	0.647	0.374	0.464	0.625	0.607	0.288	0.586
HEOG theta	0.569	0.330	0.464	0.587	0.412	0.266	0.577
HEOG alpha	0.523	0.330	0.388	0.606	0.412	0.298	0.580
HEOG beta	0.566	0.378	0.490	0.639	0.439	0.384	0.622
HEOG gamma	0.650	0.519	0.703	1.000	0.699	0.467	0.797
VEOG delta	0.580	0.416	0.500	0.520	0.603	0.514	0.421
VEOG theta	0.547	0.437	0.434	0.611	0.591	0.326	0.482
VEOG alpha	0.554	0.386	0.361	0.535	0.567	0.358	0.406
VEOG beta	0.502	0.353	0.392	0.448	0.510	0.348	0.420
VEOG gamma	0.679	0.521	0.616	0.712	0.727	0.604	0.561
FZ delta	0.445	0.307	0.372	0.440	0.432	0.268	0.489
FZ theta	0.463	0.349	0.433	0.535	0.439	0.326	0.473
FZ alpha	0.454	0.339	0.353	0.440	0.490	0.325	0.474
FZ beta	0.591	0.395	0.442	0.597	0.584	0.375	0.514
FZ gamma	0.649	0.524	0.531	0.652	0.657	0.450	0.652
F7 delta	0.459	0.431	0.452	0.520	0.548	0.310	0.464
F7 theta	0.447	0.309	0.422	0.490	0.397	0.284	0.510
F7 alpha	0.522	0.395	0.389	0.386	0.506	0.345	0.485
F7 beta	0.533	0.518	0.523	0.562	0.390	0.572	0.528
F7 gamma	0.758	0.664	0.646	0.789	0.685	0.643	0.699
PZ delta	0.580	0.356	0.368	0.457	0.399	0.373	0.597
PZ theta	0.497	0.356	0.372	0.391	0.449	0.333	0.510
PZ alpha	0.611	0.394	0.326	0.391	0.618	0.287	0.438
PZ beta	0.476	0.419	0.388	0.506	0.523	0.391	0.472
PZ gamma	0.568	0.486	0.491	0.487	0.684	0.461	0.508
T5 delta	0.513	0.401	0.423	0.416	0.566	0.301	0.461
T5 theta	0.498	0.324	0.384	0.413	0.499	0.319	0.430
T5 alpha	0.460	0.338	0.341	0.469	0.478	0.300	0.524
T5 beta	0.623	0.549	0.644	0.632	0.664	0.523	0.719
T5 gamma	0.781	0.663	0.748	0.908	0.801	0.637	0.676
O2 delta	0.523	0.340	0.398	0.425	0.499	0.260	0.432
O2 theta	0.636	0.388	0.388	0.468	0.686	0.317	0.535
O2 alpha	0.774	0.447	0.459	0.591	0.884	0.314	0.613
O2 beta	0.650	0.416	0.513	0.611	0.651	0.409	0.561
O2 gamma	0.711	0.683	0.708	0.886	0.866	0.747	0.685
Ameter	0.442	0.391	0.540	0.447	0.405	0.515	0.522
Interbeat	0.591	0.465	0.528	0.785	0.536	0.453	0.762
Interblink	0.943	0.523	0.625	0.775	1.000	0.346	0.830
EDA Tonic	0.750	1.000	0.931	0.895	0.776	1.000	0.838
EMG	0.753	0.998	1.000	0.833	1.000	0.951	0.918
Pupil Diam	1.000	0.789	0.887	0.940	0.867	0.759	1.000

Appendix E. Sorted Features for Each Cognitive Load Grouping

Spatial Working Memory	Verbal Working Memory	Executive Function	Global Workload	Spatial vs Verbal	VHT Task	OVI Task
Pupil Diam	EDA Tonic	EMG	HEOG gamma	Interblink	EDA Tonic	Pupil Diam
Interblink	EMG	EDA Tonic	Pupil Diam	EMG	EMG	EMG
T5 gamma	Pupil Diam	Pupil Diam	T5 gamma	O2 alpha	Pupil Diam	EDA Tonic
O2 alpha	O2 gamma	T5 gamma	EDA Tonic	Pupil Diam	O2 gamma	Interblink
F7 gamma	F7 gamma	O2 gamma	O2 gamma	O2 gamma	F7 gamma	HEOG gamma
EMG	T5 gamma	HEOG gamma	EMG	T5 gamma	T5 gamma	Interbeat
EDA Tonic	T5 beta	F7 gamma	F7 gamma	EDA Tonic	VEOG gamma	T5 beta
O2 gamma	FZ gamma	T5 beta	Interbeat	VEOG gamma	F7 beta	F7 gamma
VEOG gamma	Interblink	Interblink	Interblink	HEOG gamma	T5 beta	O2 gamma
HEOG gamma	VEOG gamma	VEOG gamma	VEOG gamma	O2 theta	Ameter	T5 gamma
O2 beta	HEOG gamma	Ameter	FZ gamma	F7 gamma	VEOG delta	FZ gamma
FZ gamma	F7 beta	FZ gamma	HEOG beta	PZ gamma	HEOG gamma	HEOG beta
HEOG delta	PZ gamma	Interbeat	T5 beta	T5 beta	PZ gamma	O2 alpha
O2 theta	Interbeat	F7 beta	HEOG delta	FZ gamma	Interbeat	PZ delta
T5 beta	O2 alpha	O2 beta	O2 beta	O2 beta	FZ gamma	HEOG delta
PZ alpha	VEOG theta	VEOG delta	VEOG theta	PZ alpha	O2 beta	HEOG alpha
FZ beta	F7 delta	PZ gamma	HEOG alpha	HEOG delta	PZ beta	HEOG theta
Interbeat	PZ beta	HEOG beta	FZ beta	VEOG delta	HEOG beta	VEOG gamma
VEOG delta	VEOG delta	HEOG delta	O2 alpha	VEOG theta	FZ beta	O2 beta
PZ delta	O2 beta	HEOG theta	HEOG theta	FZ beta	PZ delta	O2 theta
HEOG theta	T5 delta	O2 alpha	F7 beta	VEOG alpha	VEOG alpha	F7 beta
PZ gamma	FZ beta	F7 delta	VEOG alpha	T5 delta	VEOG beta	T5 alpha
HEOG beta	F7 alpha	FZ beta	FZ theta	F7 delta	Interblink	Ameter
VEOG alpha	PZ alpha	VEOG theta	VEOG delta	Interbeat	F7 alpha	FZ beta
VEOG theta	Ameter	FZ theta	F7 delta	PZ beta	PZ theta	PZ theta
F7 beta	O2 theta	T5 delta	PZ beta	VEOG beta	VEOG theta	F7 theta
HEOG alpha	VEOG alpha	F7 theta	F7 theta	F7 alpha	FZ theta	PZ gamma
O2 delta	HEOG beta	O2 delta	PZ gamma	O2 delta	FZ alpha	FZ delta
F7 alpha	HEOG delta	VEOG beta	T5 alpha	T5 theta	T5 theta	F7 alpha
T5 delta	PZ theta	F7 alpha	O2 theta	FZ alpha	O2 theta	VEOG theta
VEOG beta	PZ delta	PZ beta	PZ delta	T5 alpha	O2 alpha	FZ alpha
T5 theta	VEOG beta	O2 theta	VEOG beta	PZ theta	F7 delta	FZ theta
PZ theta	FZ theta	HEOG alpha	Ameter	HEOG beta	T5 delta	PZ beta
PZ beta	O2 delta	T5 theta	FZ alpha	FZ theta	T5 alpha	F7 delta
FZ theta	FZ alpha	FZ delta	FZ delta	FZ delta	HEOG alpha	T5 delta
T5 alpha	T5 alpha	PZ theta	O2 delta	HEOG theta	HEOG delta	PZ alpha
F7 delta	HEOG alpha	PZ delta	T5 delta	HEOG alpha	PZ alpha	O2 delta
FZ alpha	HEOG theta	VEOG alpha	T5 theta	Ameter	F7 theta	T5 theta
F7 theta	T5 theta	FZ alpha	PZ alpha	PZ delta	FZ delta	VEOG delta
FZ delta	F7 theta	T5 alpha	PZ theta	F7 theta	HEOG theta	VEOG beta
Ameter	FZ delta	PZ alpha	F7 alpha	F7 beta	O2 delta	VEOG alpha

Appendix F. Confusion Matrices for Cognitive Gauges during the Single Task Experiments

Spatial Working Memory

Testing Probability Matrix *				
	No Spatial	Low Spatial	High Spatial	Total
No Spatial	90.36	8.67	0.97	48.61
Low Spatial	13.26	72.53	14.20	24.12
High Spatial	2.39	25.35	72.26	27.27
Total	47.77	28.62	23.61	100.00

Verbal Working Memory

Testing Probability Matrix *				
	No Verbal	Low Verbal	High Verbal	Total
No Verbal	93.29	3.32	3.39	49.28
Low Verbal	10.97	59.08	29.95	25.40
High Verbal	15.20	43.81	40.99	25.32
Total	52.60	27.74	19.66	100.00

Executive Function

Testing Probability Matrix *				
	Low Executive	Med Executive	High Executive	Total
Low Executive	84.32	13.36	2.32	58.10
Med Executive	46.02	51.40	2.58	26.46
High Executive	45.78	4.42	49.80	15.44
Total	68.23	22.05	9.72	100.00

Global Workload

Testing Probability Matrix *			
	Low Global	High Global	Total
Low Global	79.83	20.17	58.07
High Global	46.35	53.65	41.93
Total	65.79	34.21	100.00

*** Rows indicate the actual class and columns represent predicted class. All numbers are percent assigned to group**

Spatial versus Verbal Working Memory

Testing Probability Matrix *			
	Verbal	Spatial	Total
Verbal	89.49	10.51	48.61
Spatial	7.14	92.86	51.39
Total	47.16	52.84	100.00

VHT

Testing Probability Matrix *			
	Low VHT	High VHT	Total
Low VHT	62.70	37.30	50.12
High VHT	44.64	55.36	49.88
Total	53.69	46.31	100.00

OVI Task

Testing Probability Matrix *				
	Cruise	Low SAR	High SAR	Total
Cruise	82.53	11.82	5.66	46.93
Low SAR	26.63	61.59	11.78	23.04
High SAR	26.42	15.52	58.06	30.03
Total	52.80	24.40	22.80	100.00

*** Rows indicate the actual class and columns represent predicted class. All numbers are percent assigned to group**

Appendix G. Raw Data for Hit, Miss, False Alarm, and Correct Rejection for Each SAR Image Grouped by Aiding Type and Workload Level

Class	Hit	Miss	False Alarm	Correct Rejection
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	5	1	1	6
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	4	2	2	5
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	4	2	2	5
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	6	0	0	6

Class	Hit	Miss	False Alarm	Correct Rejection
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	6	0	0	6
High Spatial - No Aiding	0	6	0	6
High Spatial - No Aiding	0	6	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	0	6	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	0	6	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	6	0	1	5
High Spatial - Aiding	6	0	1	5
High Spatial - Aiding	5	1	1	5
High Spatial - Aiding	5	1	1	5
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	6	0	0	6
High Spatial - Aiding	5	1	1	5
High Spatial - Random Aiding	6	0	0	6
High Spatial - Random Aiding	6	0	0	6
High Spatial - Random Aiding	0	6	0	6
High Spatial - Random Aiding	6	0	0	6
High Spatial - Random Aiding	6	0	0	6
High Spatial - Random Aiding	6	0	0	6
High Spatial - Random Aiding	0	6	0	6
High Spatial - Random Aiding	6	0	0	6
High Spatial - Random Aiding	5	1	2	4
High Spatial - Random Aiding	6	0	0	6
High Spatial - Random Aiding	0	6	0	6
High Spatial - Random Aiding	0	6	0	6
High Spatial - Random Aiding	6	0	0	6
High Spatial - Random Aiding	6	0	0	6
High Spatial - Random Aiding	6	0	0	6

Class	Hit	Miss	False Alarm	Correct Rejection
High Spatial - Random Aiding	6	0	0	6
High Spatial - Random Aiding	5	1	2	4
High Spatial - Random Aiding	4	2	2	4
High Spatial - Random Aiding	5	1	1	5
High Spatial - Random Aiding	0	6	0	6
High Spatial - Random Aiding	0	6	0	6
High Spatial - Random Aiding	6	0	0	6
High Spatial - Random Aiding	6	0	0	6
High Spatial - Random Aiding	0	6	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	5	1	1	5
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6

Class	Hit	Miss	False Alarm	Correct Rejection
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	0	6	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	6	0	0	6
Low Spatial - No Aiding	0	6	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6

Class	Hit	Miss	False Alarm	Correct Rejection
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	5	1	1	5
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6
Low Spatial - Random Aiding	6	0	0	6

Appendix H. Raw Data for Missed Weapons Release Waypoint and Number of DMPIs Placed for Each SAR Image Grouped by Aiding Type and Workload Level

Class	Missed Weapons Release	DMPIs Placed
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	1	5
High Spatial - No Aiding	0	6
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	1	0
High Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6

Class	Missed Weapons Release	DMPIs Placed
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	1	2
High Spatial - No Aiding	1	4
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
High Spatial - Aiding	1	4
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	5
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	1	2
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
High Spatial - No Aiding	0	2
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	5
High Spatial - No Aiding	1	3
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6

Class	Missed Weapons Release	DMPIs Placed
High Spatial - Aiding	1	3
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	1	1
High Spatial - Random Aiding	1	2
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	1	2
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	1	5
High Spatial - No Aiding	0	6
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6

Class	Missed Weapons Release	DMPIs Placed
High Spatial - No Aiding	0	3
High Spatial - No Aiding	1	2
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	1	3
High Spatial - No Aiding	1	2
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	1	0
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	1	0
High Spatial - No Aiding	1	0
Low Spatial - No Aiding	1	4
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	0	6
Low Spatial - No Aiding	1	3
High Spatial - No Aiding	0	5
High Spatial - No Aiding	0	6
High Spatial - No Aiding	0	6
High Spatial - No Aiding	1	0
Low Spatial - Aiding	0	6

Class	Missed Weapons Release	DMPIs Placed
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
Low Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Aiding	0	6
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	0	5
High Spatial - Random Aiding	0	6
High Spatial - Random Aiding	1	1
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6
Low Spatial - Random Aiding	0	6

Bibliography

Alsing, Stephen G., *The Evaluation of Competing Classifiers*, Air Force Institute of Technology (AU), Wright-Patterson AFB, OH, March 2000.

Alsing, Stephen G., Kenneth W. Bauer, and John O. Miller. "A Multinomial Selection Procedure for Evaluating Pattern Recognition Algorithms," *Pattern Recognition*, 35: 2397-2412 (2002).

Anderson, Charles W., Erik A. Stolz, and Sanyogita Shamsunder. "Discriminating Mental Tasks Using EEG Represented by AR Models," *Proceedings of the 1995 IEEE Engineering in Medicine and Biology Annual Conference*, Montreal, Canada, Sep 20-23, 1995.

Anderson, Charles W., Erik A. Stolz, and Sanyogita Shamsunder. "Multivariate Autoregressive Models for Classification of Spontaneous Electroencephalographic Signals During Mental Tasks," *IEEE Transactions On Biomedical Engineering*, 45(3): 277-286 (1998).

Anderson, Charles W., Saikumar V. Devulapalli, and Erik A. Stolz. "Determining Mental State from EEG Signals Using Parallel Implementations of Neural Networks," *Scientific Programming, Special Issue on Applications Analysis*, 4(3): 171-183 (1995).

Barry, Charles L. and Elihu Zimet. "UCAVs – Technological, Policy, and Operational Challenges," *Defense Horizons*, 3: 1-8 (October 2001).

Bauer, Kenneth W., Stephen G. Alsing, and Kelly A. Greene. "Feature Screening using Signal to-Noise Ratios," *Neurocomputing*, 31: 29-44 (2000).

Beaumont, M., A. Burov, R. Carter, S.N. Cheuvront, M.N. Sawka, G. Wilson, K. Van Orden, B. Hockey, T. Balkin and A. Gundel, "Risk Factors – Individual State," in *Operator Functional State Assessment* Eds. RTO Human Factors and Medicine Panel (HFM) Task Group, RTO Technical Report TR-HFM-104, RTO/NATO, 2004.

Bishop, Christopher M. *Neural Networks for Pattern Recognition.* New York: Oxford University Press, 1995.

Borge, Jean L. "UCAV Team Finishes Phase One," *Aerospace Engineering*, 23(3):6 (2003).

Brookings, Jeffrey B., Glenn F. Wilson, and Carolyn R. Swain. "Psychophysiological Responses to Changes in Workload During Simulated Air Traffic Control," *Biological Psychology*, 42: 361-377 (1996).

Bubb-Lewis, Cristina and Mark W. Scerbo. "The Effects of Communication Modes on Performance and Discourse Organization with an Adaptive Interface," *Applied Ergonomics*, 33: 15-26 (2002).

Burges, Christopher C.J. "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, 2: 121-167 (1998).

Byrne, Evan A. and Raja Parasuraman. "Psychophysiology and Adaptive Automation," *Biological Psychology*, 42: 249-268 (1996).

Casella, George and Roger L. Berger. *Statistical Inference.* Pacific Grove, CA: Duxbury: 2002.

Costa, Ernane J.X. and Euvaldo F. Cabral Jr. "EEG-Based Discrimination Between Imagination of Left and Right Hand Movements Using Adaptive Gaussian Representation," *Medical Engineering & Physics*, 22: 345-348 (2000).

Cover, Thomas M. "Geometrical and Statistical Properties of Systems of Linear Inequalities With Applications in Pattern Recognition," *IEEE Transactions on Electronic Computers*, EC-14: 326-334 (1965).

Duda, Richard O., Peter E. Hart, and David G. Stork, *Pattern Classification.* New York: John Wiley & Sons, 2001.

Dunteman, George H. *Principle Component Analysis.* Newbury Park, CA: Sage University Press, 1989.

Flury, B. *Common Principal Components and Related Multivariate Models.* New York: John Wiley & Sons, 1988.

Fournier, Lisa R., Glenn F. Wilson, and Carolyne R. Swain. "Electrophysiological, Behavioral, and Subjective Indices of Workload When Performing Multiple Tasks: Manipulations of Task Difficulty and Training," *International Journal of Psychophysiology*, 31: 129-145 (1999).

Fraser, W., E. Svensson, M. Grandt, B. Hockey, T. Balkin, M. Beaumont, G. Kamimori, M. Kautz, G.L. Belenky, N. Wesensten and R.E. Schlegel, "Risk Factors - Environmental," in *Operator Functional State Assessment* Eds. RTO Human Factors and Medicine Panel (HFM) Task Group, RTO Technical Report TR-HFM-104, RTO/NATO, 2004.

Freeman, Frederick, G., Peter J. Mikulka, Lawrence J. Prinzel, and Mark W. Scerbo. "Evaluation of an Adaptive Automation System Using Three EEG Indices with a Visual Tracking Task," *Biological Psychology*, 50: 61-76 (1999).

Fukunaga, Keinosuke. *Introduction to Statistical Pattern Recognition, Second Edition.* San Diego, CA: Academic Press, 1990.

Gaillard, A.W.K. and A.F. Kramer. "Theoretical and Methodological Issues in Psychophysiological Research," in *Engineering Psychophysiology: Issues and Applications*. Eds. R.W. Backs and W. Boucsein, Mahwah, New Jersey: Lawrence Erlbaum Associates, 2000.

Galicki, Miroslaw, Herbert Witte, Jens Dörschel, Michael, Eiselt, and Gert Griessbach. "Common Optimization of Adaptive Preprocessing Units and a Neural Network during the Learning Period. Application in EEG Pattern Recognition," *Neural Networks*, 10(6): 1153-1163 (1997).

Garner, Major Mark E, *Human/Systems Interface (HIS) Issues from a USAF Pilot/Combat Unmanned Air Vehicle (UAV) Test Operator Perspective.* Technical Report AFRL-VA-WP-TP-202-310, Wright-Patterson AFB, OH, April 2002.

Garrett, Deon, David A. Peterson, Charles W. Anderson, and Michael H. Thaut. "Comparison of Linear, Nonlinear, and Feature Selection Methods for EEG Signal Classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2): 141-144 (2003).

Gasser, Theo, Lothar Sroka and Joachim Möcks. "The Transfer of EOG Activity into the EEG for Eyes Open and Closed," *Electroencephalography and clinical Neurophysiology*, 61: 181-193 (1985).

Gershenfeld, Neil. *The Nature of Mathematical Modeling* Cambridge, Great Britain: Cambridge University Press, 1999.

Gevins, Alan, Michael E. Smith, Harrison Leong, Linda K. McEvoy, Susan Whitfield, Robert Du, and Georgia Rush. "Monitoring Working Memory Load During Computer-based Tasks with EEG Pattern Recognition Methods," *Human Factors*, 40(1): 79-91 (1998).

Gevins, Alan S. and Nelson H. Morgan. "Classifier-Directed Signal Processing in Brain Research," *IEEE Transactions on Biomedical Engineering*, BME-33(12): 1054-1067 (1986).

Gevins, Alan S. and Nelson H. Morgan. "Applications of Neural-Network (NN) Signal Processing in Brain Research," *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(7): 1152-1161 (1988).

Gevins, Alan and Michael E. Smith. "Detecting Transient Cognitive Impairment with EEG Pattern Recognition Methods," *Aviation, Space and Environmental Medicine*, 70(10): 1018-1024 (1999).

Greene, Kelly A., Kenneth W. Bauer, Glenn F. Wilson, Chris A. Russell, Stephen K. Rogers, and Matthew Kabrisky. "Selection of Psychophysiological Features for Classifying Air Traffic Controller Workload in Neural Networks," *Smart Engineering Systems Design*, 2: 315-330 (2000).

Greene, Kelly A., Kenneth W. Bauer, Matthew Kabrisky, Stephen K. Rogers, and Glenn F. Wilson. "Estimating Pilot Workload Using Elman Recurrent Neural Networks: A Preliminary Investigation," in *Smart Engineering Systems: Neural Networks, Fuzzy Logic and Evolutionary Programming* Eds. Cihan H. Dagli, Metin Akay, Okan Ersoy, Benito R. Fernandez, and Alice Smith, New York: ASME Press, 7: 703-708, (1997).

Grözinger, Michael, Joachim Röschke, and Bert, Klöppel. "Automatic Recognition of Rapid Eye Movement (REM) Sleep By Artificial Neural Networks," *Journal of Sleep Research*, 4: 86-91 (1995).

Guevara, M.A., I. Lorenzo, C. Arce, J. Ramos, and M. Corsi-Cabrera. "Inter- and Intrahemispheric EEG Correlation During Sleep and Wakefulness," *Sleep*, 18(4): 257-265 (1995).

Guger, C., H. Ramoser, and G. Pfurtscheller. "Real-Time EEG Analysis with Subject-Specific Spatial Patterns for a Brain-Computer Interface (BCI)," *IEEE Transactions on Rehabilitation Engineering*, 8(4): 447-456 (2000).

Guger, Christop, Alois Schlögl, Christa Neuper, Dirk Walterspacher, Thomas Strein, and Gert Pfurtscheller. "Rapid Prototyping of an EEG-Based Brain-Computer Interface (BCI)," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 9(1): 49-57 (2001).

Hand, D.J., *Construction and Assessment of Classification Rules.* West Sussex, England: John Wiley& Sons, 1997.

Hart, S.G. and L.E. Staveland. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Human Mental Workload* Eds. P.A. Hancock and N. Meshkati, Amsterdam: North_Holland, (1988).

Haselsteiner, Ernst and Gert Pfurtscheller. "Using Time-Dependent Neural Networks for EEG Classification," *IEEE Transactions on Rehabilitation Engineering*, 8(4): 457-463 (2000).

Haykin, Simon. *Neural Networks: A Comprehensive Foundation.* Upper Saddle River, NJ: Prentice Hall, 1999.

Hazarika, Neep, Jean Zhu Chen, Ah Chung Tsoi, and Alex Sergejew. "Classification of EEG Signals Using the Wavelet Transform," *Signal Processing*, 59: 61-72 (1997).

Hazarika, Neep, Ah Chung Tsoi, and Alex Sergejew. "Nonlinear Considerations in EEG Signal Classification," *IEEE Transactions on Signal Processing*, 45(4): 829-836 (1997).

He, P., G.F. Wilson, and C.A. Russell. "Removal of Ocular Artifacts from Electo-Encephalogram by Adaptive Filtering," *Medical & Biological Engineering & Computing*, 42: 407-412 (2004).

Ivanova, Irena, Gert Pfurtscheller, and Colin Andrew. "AI-Based Classification of Single-Trial EEG Data," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology*, 703-704 (1995).

Jasper, H. H. "The Ten-Twenty Electrode System of the International Federation," *Electroencephalography and clinical Neurophysiology*, 10: 371-375 (1958).

John, E.R., L.S. Prichep, J. Fridman, and P. Easton. "Neurometrics: Computer-Assisted Differential Diagnosis of Brain Dysfunctions," *Science*, 239: 162-169 (1988).

Jolliffe, Ian T. *Principal Component Analysis.* New York: Springer-Verlag, 1986.

Keirn, Zachary A. and Jorge I. Aunon. "A New Mode of Communication Between Man and His Surroundings," *IEEE Transactions on Biomedical Engineering*, 37(12): 1209-1214 (1990).

Keirn, Zachary A. and Jorge I. Aunon. "Man-Machine Communications Through Brain-Wave Processing," *IEEE Engineering in Medicine and Biology Magazine*, 55-57 March 1990.

Klöppel, Bert. "Neural Networks as a New Method for EEG Analysis," *Neuropsychobiology*, 29: 33-38 (1994).

Kirsch, Peter, Christoph Besthorn, Sabine Klein, Jochen Rindfleisch, and Robert Olbrich. "The Dimensional Complexity of the EEG During Cognitive Tasks Reflects the Impaired Information Processing in Schizophrenic Patients," *International Journal of Psychophysiology*, 36: 237-246 (2000).

Kudo, Terence Y., *Using Statistical Process Control Methods to Classify Pilot Mental Workload.* Air Force Institute of Technology (AU), Wright-Patterson AFB, OH, March 2001.

Kurooka, Taketoshi, Yuh Yamashita, and Hirokazu Nishitani. "Mind State Estimation for Operator Support," *Computers and Chemical Engineering*, 24: 551-556 (2000).

Laine, Trevor I., Kenneth W. Bauer, Jr., Jeffrey W. Lanning, Chris A. Russell, and Glenn F. Wilson. "Selection of Input Features Across Subjects for Classifying Crewmember Workload Using Artificial Neural Networks," *IEEE Transactions on Systems, Man, and Cybernetics – Part A. Systems and Humans*, 32(6): 691-704 (2002).

Lal, Thomas Navin, Michael Schröder, Thilo Hinterberger, Jason Weston, Martin Bogdan, Nils Birbaumer, and Bernhard Schölkopf. "Support Vector Channel Selection in BCI," *IEEE Transactions on Biomedical Engineering*, 51(6): 1003-1010 (2004).

Lippmann, R. P. "An Introduction to Computing with Neural Nets," *IEEE Acoustics, Speech, and Signal Processing Magazine*, 4(2): 4-22 (1987).

Liu, He Sheng, Tong Zhang, and Fu Sheng Yang. "A Multistage, Multimethod Approach for Automatic Detection and Classification of Epileptiform EEG," *IEEE Transactions on Biomedical Engineering*, 49(12): 1557-1566 (2002).

Lu, Bao-Liang, Jonghan Shin, and Michinori Ichikawa. "Massively Parallel Classification of Single-Trial EEG Signals Using a Min-Max Modular Neural Network," *IEEE Transactions on Biomedical Engineering*, 51(3): 551-558 (2004).

Makeig, Scott and Tzyy-Ping Jung. "Tonic, Phasic, and Transient EEG Correlates of Auditory Awareness in Drowsiness," *Cognitive Brain Research*, 4: 15-25 (1996).

Makeig, Scott, Anthony J. Bell, Tzyy-Ping Jung, and Terrence J. Sejnowski. "Independent Component Analysis of Electroencephalographic Data," in *Advances in Neural Information Processing Systems.* Eds. D. Touretzky, M. Mozer, and M. Hasselmo. Cambridge, MA: MIT Press, 8: 145-151 (1996).

Makeig, Scott, Tzyy-Ping Jung, and Terrence J. Sejnowski. "Using Feedforward Networks to Monitor Alertness from Changes in EEG Correlation and Coherence," in *Advances in Neural Information Processing Systems.* Eds. D. Touretzky, M. Mozer, and M. Hasselmo. Cambridge, MA: MIT Press, 8: 931-937 (1996).

Makeig, Scott, Sigurd Enghoff, Tzyy-Ping Jung, and Terrence J. Sejnowski. "A Natural Basis for Efficient Brain-Actuated Control," *IEEE Transactions on Rehabilitation Engineering*, 8(2): 208-211 (2000).

Marshall, Sandra, Conversations with Dr. Sandra Marshall of Eyetracking, Inc., San Diego, CA. and San Diego State University during the DARPA Augmented Cognition Concept Validation Experiment in Seattle, WA., 29 March 2004 – 1 April 2004.

Mason, S.G. and G.E. Birch. "A Brain-Controlled Switch for Asynchronous Control Applications," *IEEE Transactions on Biomedical Engineering*, 47(10): 1297-1306 (2000).

McEvoy, L. K., M. E. Smith, and A. Gevins. "Test-Retest Reliability of Cognitive EEG," *Clinical Neurophysiology*, 111: 457-463 (2000).

Millán, José del R., Josep Mouriño, Marco Franzé, Febo Cincotti, Markus Varsta, Jukka Heikkonen, and Fabio Babiloni. "A Local Neural Classifier for the Recognition of EEG Patterns Associated to Mental Tasks," *IEEE Transactions on Neural Networks*, 13(3): 678-686 (2002).

Müller, Klaus-Robert, Sebastian Mika, Gunnar Rätsch, Koji Tsuda, and Bernhard Schölkopf. "An Introduction to Kernel-Based Learning Algorithms," *IEEE Transactions on Neural Networks*, 12(2): 181-202 (2001).

Müller, Klaus-Robert, Charles W. Anderson, and Gary E. Birch. "Linear and Nonlinear Methods for Brain-Computer Interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2): 165-169 (2003).

Müller-Gerking, Johannes, Gert Pfurtscheller, and Henrik Flyvbjerg. "Classification of Movement-Related EEG in a Memorized Delay Task Experiment," *Clinical Neurophysiology*, 111: 1353-1365 (2000).

Neter, John, Michael H. Kutner, Christopher J. Nachtsheim, and William Wasserman. *Applied Linear Statistical Models, Fourth Edition.* Boston, MA: WCB McGraw-Hill, 1996.

Obermaier, Bernhard, Christa Neuper, Christop Guger, and Gert Pfurtscheller. "Information Transfer Rate in a Five-Classes Brain-Computer Interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 9(3): 283-288 (2001).

Obermaier, B., C. Guger, C. Neuper, and G. Pfurtscheller. "Hidden Markov Models for Online Classification of Single Trial EEG Data," *Pattern Recognition Letters*, 22: 1299-1309 (2001).

Parasuraman, Raja. "Adaptive Automation Matched to Human Mental Workload," in *Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks.* Eds. G. Robert J. Hockey, Anthony W. K. Gaillard, and Oleksandr Burov. Amsterdam: IOS Press, 2003.

Parasuraman, Raja, Thomas B. Sheridan, and Christopher D. Wickens. "A Model for Types and Levels of Human Interaction with Automation," *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 30(3): 286-297 (2000).

Parasuraman, Raja, Mustapha Mouloua, and Robert Molloy. "Effects of Adaptive Task Allocation on Monitoring of Automated Systems," *Human Factors*, 38(4): 665-679 (1996).

Parasuraman, Raja. "Humans and Automation: Use, Misuse, Disuse, Abuse," *Human Factors*, 39(2): 230-253 (1997).

Peters, B.O., G. Pfurtscheller, and H. Flyvbjerg. "Automatic Differentiation of Multichannel EEG Signals," *IEEE Transactions on Biomedical Engineering*, 48(1): 111-116 (2001).

Peters, Björn O., Gert Pfurtscheller, and Henrik Flyvbjerg. "Mining Multi-Channel EEG For Its Information Content: An ANN-Based Method for a Brain-Computer Interface," *Neural Networks*, 11: 1429-1433 (1998).

Petrosian, A.A., D.V. Prokhorov, W. Lajara-Nanson, and R.B. Schiffer. "Recurrent Neural Network-Based Approach For Early Recognition of Alzheimer's Disease in EEG," *Clinical Neurophysiology*, 112: 1378-1387 (2001).

Pfurtscheller, Gert, Christa Neuper, Alois Schlögl, and Klaus Lugger. "Separability of EEG Signals Recorded During Right and Left Motor Imagery Using Adaptive Autoregressive Parameters," *IEEE Transactions on Rehabilitation Engineering*, 6(3): 316-325 (1998).

Pleydell-Pearce, Christopher, W., Sharron E. Whitecross, and Blair T. Dickson. "Multivariate Analysis of EEG: Predicting Cognition on the Basis of Frequency Decomposition, Inter-electrode Correlation, Coherence, Cross Phase and Cross Power," *Proceedings of the 36th Hawaiian International Conference of System Sciences*, 131-141, 2003.

Polak, Mark and Aleksander Kostov. "Development of Brain-Computer Interface: Preliminary Results," *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology*, 1543-1546 (1997).

Pregenzer, Martin and Gert Pfurtscheller. "Frequency Component Selection for an EEG-Based Brain to Computer Interface," *IEEE Transactions on Rehabilitation Engineering*, 7(4): 413-419 (1999).

Prinzel, Lawrence J., Frederick G. Freeman, Mark W. Scerbo, Peter J. Mikulka, and Alan T. Pope. "A Closed-Loop System for Examining Psychophysiological Measures for Adaptive Task Allocation," *The International Journal of Aviation Psychology*, 10(4): 393-410 (1999).

Pritchard, Walter S., Dennis W. Duke, Kerry L. Coburn, Norman C. Moore, Karen A. Tucker, Michael W. Jann, and Russell M. Hostetler. "EEG-Based, Neural-Net Predictive Classification of Alzheimer's Disease Verses Control Subjects Is Augmented By Non-Linear EEG Measures," *Electroencephalography and clinical Neurophysiology*, 91: 118-130 (1994).

Pucci, E., N. Belardinelli, G. Cacchiò, M. Signorino, and F. Angeleri. "EEG Power Spectrum Differences in Early and Late Onset Forms of Alzheimer's Disease," *Clinical Neurophysiology*, 110: 621-631 (1999).

Ramoser, Herbert, Johannes Müller-Gerking, and Gert Pfurtscheller. "Optimal Spatial Filtering of Single Trial EEG During Imagined Hand Movement," *IEEE Transactions on Rehabilitation Engineering*, 8(4): 441-446 (2000).

Razoumnikova, Olga M. "Functional Organization of Different Brain Areas During Convergent and Divergent Thinking: An EEG Investigation," *Cognitive Brain Research*, 10: 11-18 (2000).

Raudys, Š. "How Good are Support Vector Machines?" *Neural Networks*, 13: 17-19 (2000).

Ripley, B.D. *Pattern Recognition and Neural Networks*. Cambridge, Great Britain: University Press, 1996.

Robert, Claude, Jean-François Gaudy, and Aimé Limoge. "Electroencephalogram Processing using Neural Networks," *Clinical Neurophysiology*, 113: 694-701 (2002).

Roberts, S. and L. Tarassenko. "Analysis of the Sleep EEG Using a Multilayer Network with Spatial Organisation," *IEE Proceedings-F*, 139(6): 420-425 (1992).

Robertson, Colin and Jacob Empson. "Slowed Cognitive Processing and High Workload in Parkinson's Disease," *Journal of the Neurological Sciences*, 162: 27-33 (1999).

Rosner, Bernard. *Fundamentals of Biostatistics*, Belmont, CA: Wadsworth, 1995.

RTO Human Factors and Medicine Panel Task Group. *Operator Functional State Assessment*. RTO Technical Report TR-HFM-104, Research and Technology Organisation, North Atlantic Treaty Organisation, BP 25, F-92201 Neuilly-sur-Seine Cedex, France, February 2004.

Ruck, Dennis W., Steven, K. Rogers, and Matthew Kabrisky. "Feature Selection using a Multiplayer Perceptron," *Journal of Neural Network Computing*, 2(2): 40-48 (1990).

Russell, Chris A., Corrina T. Monett, and Glenn F. Wilson. "Mental Workload Classification Using a Backpropagation Neural Network," in *Smart Engineering Systems: Neural Networks, Fuzzy Logic and Evolutionary Programming* Eds. Cihan H. Dagli, Metin Akay, C.L. Phillip Chin, Benito R. Fernandez, and Joydeep Ghosh, New York: ASME Press, 6: 685-690 (1996).

Russell, Chris A. and Glenn F. Wilson. "Air Traffic Controller Functional State Classification Using Neural Networks," in *Smart Engineering Systems: Neural Networks, Fuzzy Logic and Evolutionary Programming* Eds. Cihan H. Dagli, Metin Akay, Anna L. Buczak, Okan Ersoy, and Benito Fernandez, New York: ASME Press, 8: 649-654, (1998).

Russell, Chris A. and Grant R. McMillan. "Facial Gesture Recognition For Alternative Control," in *Smart Engineering Systems: Neural Networks, Fuzzy Logic and Evolutionary Programming* Eds. Cihan H. Dagli, Anna L. Buczak, Joydeep Ghosh, Mark Embrechts, and Okan Ersoy, New York: ASME Press, 9: 963-968, (1999).

Russell, Chris A., Gary B. Reid, and Michael A. Vidulich. "Pilot Workload Classification Using Artificial Neural Networks in a Simulated Scud Hunt Mission," *Proceedings of the 44th Annual Meeting Human Factors and Ergonomics Society*, 3:109 (2000).

Russell, Chris A. and Steven C. Gustafson. *Selecting Salient Features of Psychophysiological Measures*. Technical Report AFRL-HE-TR-2001-0136, Wright-Patterson AFB, OH, June 2001.

Sall, John, Ann Lehman, and Lee Creighton. *JMP® Start Statistics: A Guide to Statistics and Data Analysis*. Pacific Grove, CA: SAS Institute, Inc., 2001.

Scharf, Louis L. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

Schaul, Neil. "The Fundamental Neural Mechanisms of Electroencephalography," *Electroencephalography and clinical Neurophysiology*, 106: 101-107 (1998).

Scheaffer, Richard L. and James T. McClave. *Probability and Statistics for Engineers, Second Edition*. Boston: Duxbury Press, 1986.

Schiff, Steven J., Akram Aldroubi, Michael Unser, and Susumu Sato. "Fast Wavelet Transform of EEG," *Electroencephalography and clinical Neurophysiology*, 91: 442-455 (1994).

Schober, F., R. Schellenberg, and W. Dimpfel. "Reflection of Mental Exercise in the Dynamic Quantitative Topographical EEG," *Neuropsychobiology*, 31: 98-112 (1995).

Air Force Scientific Advisory Board, *UAV Technologies and Combat Operations*. Technical Report SAF-PA 96-1204, 1996.

Shanmugan, K.S. and A.M. Breipohl. *Random Signals: Detection, Estimation and Data Analysis*, New York: John Wiley & Sons, 1988.

Siegel, Sidney. *Nonparametric Statistics for the Behavioral Sciences.* New York, NY: WCB McGraw-Hill, 1956.

Siegel, Brian K. and Kirby J. Keller. "Pilot Task Monitoring Using Neural Networks," *Proceedings of the IEEE National Aerospace and Electronics Conference*, 709-714 (1992).

Simmons, George F. *Introduction to Topology and Modern Analysis*, New York: McGraw-Hill, 1963.

Slobounov, S.M., K. Fukada, R. Simon, M. Rearick, and W. Ray. "Neurophysiological and Behavioral Indices of Time Pressure Effects on Visuomotor Task Performance," *Cognitive Brain Research*, 9: 287-298 (2000).

Smith, Michael E., Linda K. McEvoy, and Alan Gevins. "Neurophysiological Indices of Strategy Development and Skill Acquisition," *Cognitive Brain Research*, 7: 389-404 (1999).

Smith, Michael .E., Alan Gevins, Halle Brown, Arati Karnik, and Robert Du. "Monitoring Task Loading with Multivariate EEG Measures during Complex Forms of Human-Computer Interaction," *Human Factors*, 42(3):366-380 (2001).

Schmorrow, Dylan. *DARPA AUGCOG Technical Integration Experiment*. DARPA Team Meeting, San Diego, CA, 4 -6 March 2003.

Sterman M.B. and C.A. Mann. "Concepts and Applications of EEG Analysis in Aviation Performance Evaluation," *Biological Psychology*, 40: 115-130 (1995).

Sterman, Maurice B., Christopher A. Mann, David A. Kaiser, and Brandall Y. Suyenobu. "Multiband Topograhic EEG Analysis of a Simulated Visuomotor Aviation Task," *International Journal of Psychophysiology*, 16: 49-56 (1994).

Szava, S., P. Valdes, R. Biscay, L. Galan, J. Bosch, I. Clark, and J.C. Jimenez. "High Resolution Quantitative EEG Analysis," *Brain Topography*, 6(3): 211-219 (1994).

Szczuka, Marcin and Piotr Wojdyłło. "Neuro-Wavelet Classifiers for EEG Signals Based on Rough Set Methods," *Neurocomputing*, 36: 103-122 (2001).

Uchida, Sunao, Irwin Feinberg, Jonathan D. March, Yoshikata Atsumi, and Tom Maloney. "A Comparison of Period Amplitude Analysis and FFT Power Spectral Analysis of All-Night Human Sleep EEG," *Physiology and Behavior*, 67(1): 121-131 (1999).

Valdés, P., J. Bosch, R. Grave, J. Hernandez, J. Riera, R. Pascual, and R. Biscay. "Frequency Domain Models of the EEG," *Brain Topography*, 4(4): 309-319 (1992).

Van Boxtel, A., W. Waterink, and I.J.T. Veldhuizen. "Tonic Facial EMG Activity As An Index of Mental Effort: Effects of Work Rate, Time-On-Task, and Achievement Motivation," *Journal of Psychophysiology*, 11: 91 (1997).

Vapnik, Vladimir N. *The Nature of Statistical Learning Theory, Second Edition*, New York: Springer-Verlag, 1999.

Vaughan, Theresa M., Laurie A. Miner, Dennis J. McFarland, and Jonathan R. Wolpaw. "EEG-Based Communication: Analysis of Concurrent EMG Activity," *Electroencephalography and clinical Neurophysiology*, 107: 428-433 (1998).

Veldhuizen, I.J.T, A.W.K. Gaillard and J. de Vries. "The Influence of Mental Fatigue on Facial EMG Activity During a Simulated Workday," *Biological Psychology*, 63: 59-78 (2003).

Vidulich, Michael A., Conversations with Dr. Michael Vidulich of the Air Force Research Laboratory, Wright-Patterson AFB, OH, 20 August 2004.

Vuckovic, Aleksandra, Vlada Radivojevic, Andrew C.N. Chen and Dejan Popovic. "Automatic Recognition of Alertness and Drowsiness from EEG by an Artificial Neural Network," *Medical Engineering & Physics*, 24: 349–360 (2002).

Waldstein, Shari R., Willem J. Kop, Louis A. Schmidt, Amy J. Haufler, David S. Krantz, and Nathan A. Fox. "Frontal Electrocortical and Cardiovascular Reactivity During Happiness and Anger," *Biological Psychology*, 55: 3-23 (2000).

Wickens, Christopher D. *Engineering Psychology and Human Performance*, Columbus, OH: Charles E. Merrill Publishing Company, 1984.

Widrow, B. and S.D. Stearns. *Adaptive Signal Processing*, Reading, MA: Addison-Wesley, 1990.

Widrow, Bernard and Michael A. Lehr. "30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation," *Proceedings of the IEEE*, 78(9): 1415-1442 (1990).

Wilson, Glenn F., Conversations with Dr. Glenn Wilson of the Air Force Research Laboratory, Wright-Patterson AFB, OH, November 2003.

Wilson, G.F., P.S. Fullenkamp and I. Davis. "Evoked Potential, Cardiac, Blink, and Respiration Measures of Pilot Workload in Air-To-Ground Missions," *Aviation, Space, and Environmental Medicine*, 65: 100-105 (1994).

Wilson, G.F. and F. Fisher. "The Use of Cardiac and Eyeblink To Determine Flight Segment In F4 Crews," *Aviation, Space, and Environmental Medicine*, 62: 959-962 (1991).

Wilson, Glenn F., and F. Thomas Eggemeier. "Physiological Measures of Workload in Multi-Task Environments" in *Multiple-Task Performance*, Ed D. Damos, London: Taylor and Francis, 329-360 (1991).

Wilson, Glenn F., Carolyn R. Swain, and Jeffrey B. Brookings. *Workload Related Changes in Eye, Cardiac, Respiratory and Brain Activity During Simulated Air Traffic Control*. Technical Report AF/CF-TR-1995-0156, Wright-Patterson AFB, OH, August 1995.

Wilson, Glenn F., Carolyne R. Swain, and Peter Ullsperger. "EEG Power Changes During a Multiple Level Memory Retention Task," *International Journal of Psychophysiology*, 32: 107-118 (1999).

Wilson, Glenn F. and Chris A. Russell. "Heart Beat Detection Algorithm," *Unpublished Research* (2003).

Wilson, Glenn F. and Chris A. Russell. "Eyeblink Detection Algorithm," *Unpublished Research* (2003).

Wilson, Glenn F. "Pilot Workload, Operator Functional State and Adaptive Aiding," in *Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks.* Eds. G. Robert J. Hockey, Anthony W. K. Gaillard, and Oleksandr Burov. Amsterdam: IOS Press, 2003.

Wilson, Glenn F. Lambert, Jared D., and Chris A. Russell. *Performance Enhancement with Real-Time Physiologically Controlled Adaptive Aiding*. Technical Report AFRL-HE-WP-SR-2001-007, Wright-Patterson AFB, OH, August 2000.

Wilson, Glenn F. and Chris A. Russell. "Improved Performance via Real-Time Physiologically Controlled Adaptive Aiding," *Human Factors*, 45(4): 635-643 (2003).

Wilson, Glenn F. and Chris A. Russell. "Operator Functional State Classification Using Psychophysiological Features in an Air Traffic Control Task," *Human Factors*, 45(3): 381-389 (2003).

Winterer, G., B. Klöppel, A. Heinz, M. Ziller, I. G. Schmidt, and W. M. Herrmann. "Quantitative EEG (QEEG) Predicts Relapse in Patients with Chronic Alcoholism: A Combined Statistical and Neural Network Approach," *Medical and Biological Engineering and Computing*, 34(1): 235-236 (1996).

Vita

Chris A. Russell graduated from Tri-Valley High School, Dresden, Ohio in 1977. He enlisted in the United States Air Force in 1977 and graduated from basic training in 1978. He left the US Air Force after serving 12 ½ years, entered the USAF Reserves, and was assigned to the 906th Fighter Squadron. He was commissioned as a Bioenvironmental Engineer in 1993 with the 445th Air Wing where he continues to serve as a Medical Services Corps officer.

He joined the Air Force Research Laboratory in 1991 as an electrical engineer. He initially worked in the development of simulators and helmet-mounted visual display systems. He began working in the area of operator functional state estimation in 1995, developing pattern classification techniques for use in this area. A senior member of the IEEE and a life member of Tau Beta Pi, he currently has over 30 publications in the areas of pattern classification, operator functional state estimation, and digital signal processing.

He graduated from Wright State University with a BS in Electrical Engineering with honors in 1989. He continued his education, receiving an MS in Electrical Engineering in 1997. In the fall of 2000, he entered the Graduate School of Engineering and Management, Air Force Institute of Technology to pursue a Ph.D. in Electrical Engineering.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
26-08-2005	Doctoral Dissertation	Jul 2003 - Aug 2005

4. TITLE AND SUBTITLE
Operator State Estimation for Adaptive Aiding in Uninhabited Combat Air Vehicles

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Russell, Christopher A. DR-II, AFRL

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB, OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/DS/ENG/05-01

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Glenn F. Wilson, PhD
Senior Research Psychologist
AFRL/HECP
2255 H Street
WPAFB, OH 45433-7022

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT Purpose

This research demonstrated the first closed-loop implementation of adaptive automation using operator functional state in an operationally relevant environment. In the Uninhabited Combat Air Vehicle (UCAV) environment, operators can become cognitively overloaded and their performance may decrease during mission critical events. This research demonstrates an unprecedented closed-loop system, one that adaptively aids UCAV operators based on their cognitive functional state A series of experiments were conducted to 1) determine the best classifiers for estimating operator functional state, 2) determine if physiological measures can be used to develop multiple cognitive models based on information processing demands and task type, 3) determine the salient psychophysiological measures in operator functional state, and 4) demonstrate the benefits of intelligent adaptive aiding using operator functional state. Aiding the operator actually improved performance and increased mission effectiveness by 67%.

15. SUBJECT TERMS
Adaptive Automation, Uninhabited Combat Air Vehicle, Pattern Classification, Workload

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Steven C. Gustafson (ENG)
U	U	U	UU	198	19b. TELEPHONE NUMBER *(include area code)* (937)255-3636 ext 4598 Steven.gustafson@afit.edu