Dorneich, M.D., Mathan, S., Ververs, P.M, and M.C., Whitlow, S.D. (2008), "Cognitive State Estimation in Mobile Environments", *Augmented Cognition: A Practitioner's Guide*, Schmorrow D., Stanney K., & Reeves, L. (eds), Santa Monica: HFES Press.

Chapter

4

Cognitive State Estimation in Mobile Environments

Michael C. Dorneich, Santosh Mathan, Patricia May Ververs, and Stephen D. Whitlow
Honeywell Laboratories

This chapter presents the unique challenges encountered, difficult trade-offs required, and promising techniques employed when moving cognitive state assessment from the laboratory to a mobile field environment.

Introduction

Work in the field of augmented cognition began by classifying aspects of cognitive processing (attention, working memory, executive function, and sensory processing) with well-defined, well-understood laboratory tasks (often referred to informally as "Psych 101" tasks). As researchers have moved from the laboratory environment to the field environment, they have introduced the artifacts (motion, electrical, networking traffic, and disconnect) and stressors (information overload, physical load, competition, and threat of pain) inherent in some operational environments to which augmented cognition systems would be transitioned. The move from the laboratory to mobile field environments brings a number of unique challenges that must be addressed if cognitive state assessment is to be used successfully in task domains that require the operator to be mobile. Tough sacrifices need to be made, with limitations on the sensors to be used, processing power, and knowledge of the task environment. Therefore, unique techniques must be developed to enable this technology to move beyond sedentary operator domains.

Adaptive Automation

Adaptive automation, in which the automation adapts to the current task environment during task execution, either can make a certain component of a task simpler or can aid with adaptive task allocation, shifting a task from a larger multitask context to automation (Parasuraman, Mouloua, & Hilburn, 1999). Adaptive systems must make timely decisions on how best to use varying levels of automation to provide support to humans. For an adaptive system to decide when to intervene, it must have some model of the context of operations, be it a functional model of system performance or possibly a model of the operator's functional state.

Many adaptive systems derive their inferences about the cognitive state of the operator from mental models, performance on the task, or external factors related directly to the task environment (Wickens & Hollands, 2000). For example, those working in the 1980s in the field of associate systems developed adaptive information and automation management technologies that depended on a common understanding (between the automation and the human operator) of the mission, the current state of the world, the platform, and the state of the operator him- or herself. Associate systems then used that shared knowledge to plan and suggest courses of action and to adapt information displays and the behavior of automation to better serve the inferred operator intent and needs (Miller & Dorneich, 2006). Associate systems were developed for numerous domains, including single-seat fighter aircraft (Banks & Lizza, 1991), attack/scout helicopter operations (Robertson, 2000), petrochemical plants (Cochran, Miller & Bullemer, 1996), and in-home monitoring and caregiving for the elderly (Miller, Wu, Kirchbaum, & Kiff, 2004). In contrast to augmented cognition adaptive systems, the primary means used in associate systems to infer operator intent was logical deduction based on knowledge of the mission plan and the functional capabilities of the platform (Geddes, 1985).

As Scerbo et al. (2001) noted in their comparison of various adaptive automation techniques, task performance and operator modeling have advantages and disadvantages when used to drive adaptive systems. Although measurement of performance on the task has the advantage of being an online technique that can respond to unpredictable changes in the cognitive state of the operator, the method is only as good as the ability to measure performance. The use of behavioral responses to track cognitive function requires regular or periodic performance assessments to keep an updated assessment of performance capabilities. Few systems provide opportunities to track overt responses for monitoring operator performance (Parasuraman, 2003). Additionally, diagnosing cognitive state degradation via human performance degradation occurs after the fact, and thus an opportunity to proactively adapt the system to maintain performance is limited.

Modeling techniques have the advantages of offline implementation and ease of incorporation into rule-based expert systems. However, these techniques are only as good as the underlying models, and they are susceptible to model *brittleness.* Brittleness occurs because the systems model is necessarily an incomplete representation of the world (i.e., the system model does not account for all possible scenarios), and therefore the system could produce a dramatically incorrect solution when an important, but unmodeled, feature of the problem affects the choice of optimal solution (Smith, McCoy, & Layton, 1997). The more complex the task, the greater the likelihood is that the model will not anticipate all aspects of human operator performance.

Valuable Information

Augmented cognition technologies drive system adaptations by using physiological responses and brain activity to infer the availability of cognitive resources to cope with mission-relevant task demands. The goal is to enhance human performance when task-related demands surpass the human operator's assessed current cognitive capacity, which fluctuates subject to fatigue, stress, overload, or boredom. Neurophysiologically and physiologically triggered adaptive automation offers many advantages over the more traditional model-based approaches to automation by basing estimates of operator state directly on sensed data. These systems hold the promise of leveraging the strengths of humans and machines, augmenting human performance with automation specifically when assessed human cognitive capacity falls short of the demands imposed by task environments. With more refined estimates of the operator's cognitive state, measured in real time, adaptive automation also offers the opportunity to provide aid before the operator even knows he or she needs it.

The potential applications of augmented cognition cover a wide range of human-computer joint cognitive systems. One such application would be a closed-loop adaptive system to help optimize the performance of a stationary operator. Such systems may include operators who interact with information displays, such as an unmanned air vehicle ground control station operator (Snow, Barker, O'Neill, Offer, & Edwards, 2006), or an operator of a weapon control system such as the Tactical Tomahawk (Tremoulet et al. 2006).

Augmented cognition technologies can also be used for studying skill acquisition during training. Krebs et al. (1998) used positron emission tomography (PET) scans to study the learning progression of a novice operating a telerobotic arm. Cognitive state assessments during training could also be used to diagnose student difficulties in real time and provide appropriate context-specific assistance (Mathan & Dorneich, 2005). However, moving these technologies to mobile contexts remains a challenge if they are to be used in operational environments.

Knowledge of instantaneous cognitive state can be used to drive adaptive systems in many mobile contexts. Examples include pilots, dismounted soldiers, and ground vehicle operators (Dorneich, Ververs, Mathan, & Whitlow, 2005; Schnell et al., 2006, Snow et al., 2006). A truly adaptive system that manages information flow will require the ability to operate in a dynamic operational situation with a high degree of fidelity in cognitive state assessment and temporal resolution.

Physiological Measures of Cognitive State

Neurophysiologically and physiologically based assessments of cognitive state have been captured in several different ways, including electrocardiogram (ECG), electroencephalogram (EEG), and functional near-infrared (fNIR) imaging (see Chapters 1 and 2). ECG measures include heart-rate variability (HRV) in the time domain to assess mental load (Kalsbeek & Ettema, 1963), tonic heart rate to evaluate the impact of continuous information processing (Wildervanck, Mulder, & Michon, 1978), variability in the spectral domain as an index of cognitive workload (Wilson & Eggemeier, 1991), and T-wave amplitude during math interruption task performance (Heslegrave & Furedy, 1979). As discussed in Chapter 2, with fNIR spectroscopy one can conduct functional brain studies using wavelengths of light introduced at the scalp to measure cognition-related hemodynamic changes and to assess cognitive state (Izzetoglu & Bunce, 2004).

Other physiological measures used to assess cognitive state are galvanic skin response (Verwey & Veltman, 1996), eyelid movement (Neumann, 2002; Stern, Boyer, & Schroeder, 1994; Veltman & Gaillard, 1998; Yamada, 1998), pupil response (Beatty, 1982; Partala & Surakka, 2003), and respiratory patterns (Backs & Seljos, 1994; Boiten 1998; Porges & Byrne, 1992; Veltman & Gaillard, 1998; Wientjes, 1992). For a more complete review of physiological measures of mental workload, see Chapters 1 and 2, as well as Kramer (1991).

The suitability of a particular sensor to measure cognitive state depends on many factors, including ability to detect the underlying cognitive state of interest, temporal resolution needed to effectively drive mitigations, and fieldability issues in the context of use (e.g., sensor intrusiveness, processing power required, degree of operator motion).

As the "gold standard" for providing high-resolution temporal indices of cognitive activity, EEG has been used in the context of adaptive systems. Research has shown that EEG activity can be used to assess a variety of cognitive states that affect complex task performance. These include working memory (Gevins & Smith, 2000), alertness (Makeig & Jung, 1995), executive control (Garavan, Ross, Li, & Stein, 2000), and visual information processing (Thorpe, Fize, & Marlot, 1996). These findings point to the potential for using EEG measurements as the basis for driving adaptive systems that demonstrate a high degree of sensitivity and adaptability to human operators in complex task environments.

For instance, researchers have used the engagement index, developed by NASA researchers, in the context of mixed-initiative control of an automated system (Pope, Bogart, & Bartolome, 1995). This method uses a ratio of power in common frequency bands (beta / [alpha + theta]), where cognitively alert and focused are represented in beta, wakeful and relaxed in alpha, and a daydream state in theta. Higher engagement index values indicate increased levels of task engagement.

The efficacy of the engagement index as the basis for adaptive task allocation has been experimentally established. For instance, under manipulations of vigilance levels (Mikulka, Hadley, Freeman, & Scerbo, 1999) and workload (Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2000), an adaptive system effectively detected states in which human performance was likely to fall and took steps to allocate tasks in a manner that would raise overall task performance. In a different domain, adaptive scheduling of communications based on cognitive state assessment of the readiness to process information resulted in a twofold increase in message comprehension and situation awareness (Dorneich et al., 2005). These results highlight the potential benefits of a neurophysiologically triggered adaptive automation.

Challenges Inherent in Mobile Cognitive State Classification

The effectiveness of neurophysiologically triggered adaptive systems hinges on reliable and effective signal processing and cognitive state classification (see Chapter 3).

Although these are difficult technical challenges in any context, they are particularly pronounced in a system designed for mobile contexts. Assessment of an operator's state can be notably more difficult if the operator is permitted to move freely to perform cognitive tasks in conjunction with physical tasks. What is already a difficult problem—gathering clean and robust signals on which to classify cognitive state—is further complicated by signal artifacts induced by motion.

Given the potential usefulness of augmented cognition systems in mobile contexts, methods have been developed to classify cognitive state in ambulatory contexts. In this chapter, we describe the challenges inherent in mobile cognitive state classification, including the ability to (a) collect robust and clean signals, (b) create a mobile computing and data-processing infrastructure, (c) reliably classify cognitive state, and (d) experimentally assess the accuracy and specificity of the algorithms in a mobile operational setting.

The work described in this chapter was developed for the dismounted soldier—potentially one of the harshest, most mobile application domains for cognitive state estimation. Any cognitive state classification solution in this domain must be portable, efficient, and robust to extremes of conditions and motion. A robust solution that meets the challenges of this domain would result in techniques that are applicable to almost any other domain in which motion is a key component of the operator's work environment. We describe the approaches outlined in this chapter in the context of a field evaluation that tested the ability to classify cognitive workload level in an unconstrained, free-play operation with soldiers executing missions in an urban environment.

Scenario

This section presents the mobile dismounted soldier domain and the operational task context in which augmented cognition technologies were applied.

Dismounted Soldier Domain

Dismounted soldiers experience many stressors that are inherent in the operational environment, and these stressors have a direct impact on overall cognitive capabilities. For instance, physical exertion is one of the primary stressors a soldier faces on the battlefield. Simply moving to a rally point in a mission is made difficult when it requires the soldier to carry a load weighing 80 to 120 lb. (Girolamo, 2005). Other common stressors that can diminish cognition include heat (Buller, Hoyt, Ames, Latzka, & Freund, 2005; Steinman, 1987), cold, limited food and water (Buller et al., 2005; Montain, Sawka, & Wenger, 2001), fear, and sleep deprivation. Stress affects all aspects of information processing, including general arousal, selective attention, speed and accuracy of performance, and working memory (Hockey, 1986). The degradation in cognitive performance that often results from the effects of stress can have catastrophic outcomes (*APA Monitor,* 1988; U.S. Navy, 1988).

In addition to the physical stressors inherent in military operations, netcentric capabilities impose cognitive stressors on dismounted soldiers. The highly dynamic, information-rich environment of the dismounted soldier motivated the development of a toolkit for mobile classification of cognitive state. The next-generation dismounted soldier relies on netted communications to build situation awareness—the kind of situational understanding that drives decisive actions. Information exchange requirements are being pushed to the lowest levels, with the goal of enhancing the capabilities of a squad (9–10 soldiers) so that it can cover the battlefield in the same way that a platoon (16–44 soldiers) now does. The network will be characterized by a network of humans collaborating through a system of C4ISR (command, control, communications, computers, intelligence, surveillance, and reconnaissance) technologies. Small netted units will have robust team communication, state-of-the-art distributed and fused (thermal and image intensification) sensors, organic (i.e., belonging to the unit) tactical intelligence/collection assets (e.g., unmanned aerial vehicles, unmanned ground vehicles, unmanned ground sensors), and linkage to other assets to enhance situational understanding and on-the-move planning (*Future Force Warrior*, 2004). Mission success will depend on the individual soldier's ability to sort through the vast array of continuous information flow afforded by a full range of netted communications. This situation will demand the most from leaders in the field.

The increase in information flow does not come without a cost, however. Effective use of information sources is constrained by the limitations of the human cognitive system. Real-time, dynamic exchange of information in a C4ISR environment can be expected to increase the likelihood of information overload, such that postulated information superiority becomes a profound liability. Potential data overload, coupled with the efficiency of information flow required in executing military doctrine, places an over-reliance on the individual soldier. One way to ensure that soldiers are supported appropriately is to develop adaptive information management systems to promote superior situation awareness on the battlefield by assessing the soldier's readiness to receive and process information. The efficacy of such systems is contingent on reliable and timely cognitive assessment.

Domain Challenges

Soldiers are subject to extremes of motion, multiple physical and mental stressors, and a wide range of cognitive activities (long periods of vigilance punctuated by extreme periods of activity). Thus, any approach to the real-time assessment of cognitive state has to be robust to motion and noise artifacts. In addition, any cognitive state classification approach has to be robust to the potentially wide range of cognitive tasks that soldiers perform, which include simple tasks such as sentry duty or defending a position and highly complex tasks such as coordinating medical evacuations or the movements of several squads or replanning tactical moves.

General Approach and Associated Toolkit

The toolkit described here consists of four principal techniques:

1. signal collection and processing of neurophysiological and physiological signals in a mobile context;
2. classification algorithms that address individualization, bias, and generalizability;
3. a computational and experimental infrastructure to support assessment; and

4. the design of experiments to assess classification in a domain-relevant operational setting with soldiers.

Signal Collection in a Mobile Environment

Inferring cognitive state from noninvasive physiological sensors is a challenging task even in pristine laboratory environments. The signal is subject to artifacts—sensor activity that obscures or distorts information associated with the cognitive activity of interest. There is a wide range of neurophysiological and psychophysiological measures, such as EEG, ECG, PET, fNIR, functional magnetic resolution imaging (fMRI), and pupilometry, to name a few (see Chapters 1 and 2). Such measures can be used to detect and determine the cognitive state of a human user. However, only a small subset of these measures is uniquely suited for a mobile environment. Because the PET and fMRI imaging scanners are not portable, this equipment is ill-suited for mobile data collection.

The use of EEG as the basis for cognitive state assessment is motivated by characteristics such as good temporal resolution, low invasiveness, low cost, and portability. The use of ECG is motivated by the strength of the signal and maturity of ECG detection sensors. However, techniques that use EEG and ECG, as well as fNIR sensors, need to focus on ruggedizing the sensor suites and performing advanced signal processing on the data collected in order to reduce the effect of artifacts inherent with mobile participants. If the classification output is desired in real time, then designers must consider ways to efficiently perform data calculations within the time, memory, and power constraints of mobile processors. Although real-time signal processing and classification of physiological signals have been implemented previously (Berka et al., 2004; Gevins & Smith, 2003; see also Chapter 3), they have not been realized in a truly mobile, ambulatory environment.

Artifact detection and reduction, necessary to create a "clean" signal that can be classified, is driven by a consideration of the characteristics of the noise artifacts themselves. How noise artifacts are handled depends on where the noise lies in the frequency band in relation to the signal (see Figure 4.1).

Artifact reduction is driven by the noise profile.

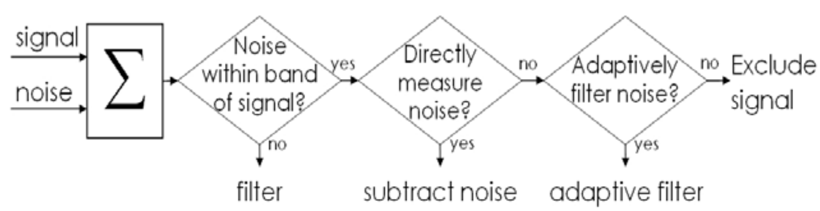


Figure 4.1. Frequency characteristics of the noise in relation to the signal of interest.

Noise signals that lie out of the band of the signal of interest can be removed with filtering. Out-of-band artifacts, such as DC drift and 60-Hz line noise, typically have well-known characteristics and can be filtered out easily. Noise artifacts that lie within the same band as the signal require more sophisticated artifact detection and reduction.

If the noise can be measured (e.g., eyeblinks measured via a dedicated ocular sensor), the sensor data can be subtracted to decontaminate the signal. However, when the noise cannot be directly measured, adaptive filtering can be applied to estimate the noise. When adaptive filtering is not feasible, the noise should be detected and the resultant data rejected, to avoid compromising downstream classification. Signal-processing approaches are specific to the signal type and dependent on factors such as signal-to-noise ratio and how specific artifacts affect the signal of interest. Two examples are discussed below: EEG and ECG.

The EEG signal is particularly subject to artifacts because of the low power in the underlying signal. High-amplitude artifacts can easily mask the lower-amplitude electrical signals associated with cognitive functions. In addition to the typical sources of signal contamination, mobile applications must consider the effects of artifacts induced by shock, cable movement, and gross muscle movement. Artifacts related to participant motion include high-frequency muscle activity, verbal communication, and ocular artifacts consisting of eye movements and blinks. Artifacts related to the operational environment include electrical noise that creates interference with the EEG signal (cf. Kramer, 1991). These concerns drive an effort to reduce the number of EEG sensors to a minimum. The minimum number of channels is dictated by the spatial resolution and underlying cognitive function of interest. More detail on these techniques is provided later in the chapter.

The challenges inherent in signal processing of ECG signals are different than those associated with EEG. Although heart rate is a very strong signal compared with an EEG signal, it is heart-rate variability that is of interest. HRV is sensitive to task demands (Aasman, Mulder, & Mulder, 1987; Beh, 1990; Porges & Raskin, 1969); thus, it is important to detect ECG peaks with a high degree of accuracy in order to identify small changes in the interbeat interval (IBI) between heart beats. In addition, it is important to account for missed peaks correctly and ignore spurious peaks, as they can have a large detrimental effect on HRV calculation.

Cognitive State Classification

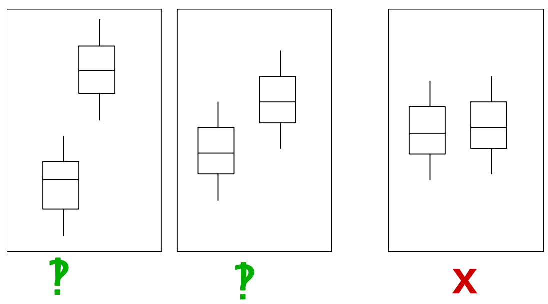
Once signal processing has been applied to create a “clean” signal, the classification stage can commence. In this section we discuss how to assess classification approaches and provide some examples of classification approaches, but this is by no means exhaustive (see also Chapter 3). Considerations are presented that help frame decisions on how to select, assess, and optimize classification.

Classification Assessment Approaches

Effective cognitive state classification approaches need to discriminate between two or more classes on a moment-to-moment basis. For discussion purposes, consider a classification algorithm that should discriminate between low and high cognitive workload—workload being an example of a cognitive state of interest. Typically, in the course of evaluating such a cognitive state classification approach, one would create a task environment containing distinct periods of high and low workload. Often, statistical tests are then conducted on the resultant data to determine if the means were statistically different in the two conditions. However, statistical significance does not suffice when determining if the classification approach is useful for cognitive state assessment. Tests of statistical significance, by definition, look at averages over an entire data set to create two distributions from the data and then determine if the means are statistically different.

Figure 4.2 shows three notional boxplot distributions between two classes. All three are statically significant given enough data points and show that the classification algorithm is tracking workload *on average*. However, it is important to know how effectively the classification approach differentiated between high and low workload on a moment-to-moment basis. The rightmost (third) distribution, though statistically different, shows considerable overlap in the distributions between the two classes. Thus, given a cognitive state classification value, it is impossible to say with confidence to which distribution it belongs.

Although differences may show up in averages, real-time classification requires an approach whereby index values in low and high conditions have minimal overlap, as in the first two distributions. all three plots may be statistically significant, but only the two at left show enough discrimination to be useful in classification between two states.



It is important to know how effectively a classification approach can differentiate between classes on a moment-to-moment basis.

Figure 4.2. Boxplot distributions.

One approach to accomplishing this is to create indices that classify workload based on each individual's unique pattern of electrophysiological activity in response to task demands. In this section, we introduce two broad approaches to classification: (a) generative classifiers that model the distribution of features in each class (examples include probability density estimation techniques such as K-nearest neighbor, Parzen windows, and Gaussian mixture models) and (b) discriminative classifiers that model a mapping function between a set of features and class labels (examples include neural nets, support vector machines, and logistic regression).

Four potential classification approaches are introduced next. The descriptions are by no means exhaustive, as many techniques have been employed, but they represent examples of both generative and discriminative approaches.

Valuable Information

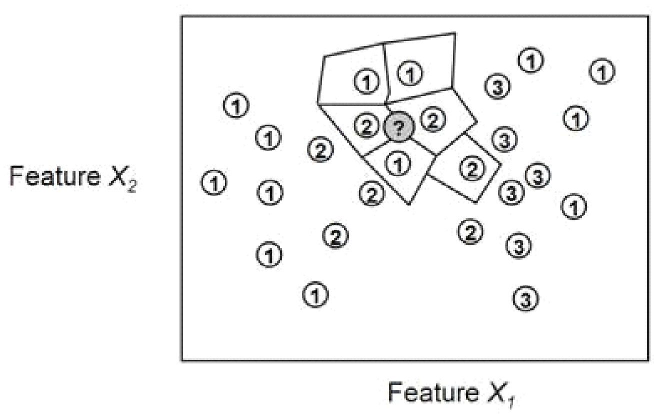
Determining a suitable classifier for a given problem is an art. Unfortunately, no single classifier approach works best on all given problems. Considerations for choosing an appropriate classifier include real-time performance, training time, and sensitivity to parameter setting. Constraints on real-time performance and training time may be dictated by the operational context. Parameter setting is important to tune a classifier to data characteristics.

Generative Classifier Approaches

Generative classification approaches have been used successfully in a mobile but constrained test task environment in which estimates of spectral power formed the input features to a pattern classification system (Mathan et al., 2005). In this example, classification systems used parametric and nonparametric techniques to assess likely cognitive state on the basis of spectral features; that is, estimate *p(cognitive state | spectral features)*. The classification process relied on probability density estimates derived from a set of spectral samples. It is important to note that when using a pattern recognition process to train the classifier, the feature set should be gathered from tasks that most closely represent the target task environment**.** Three examples of generative classification approaches are briefly introduced next: K-nearest neighbor, Parzen windows, and Gaussian mixture models.

K-nearest neighbor (KNN). The K-nearest neighbor approach is one of the simplest machine learning algorithms. It is a nonparametric technique that makes no assumption about the form of the probability densities underlying a particular set of data. Given a particular sample *x*, the classification process identifies k samples whose features come closest (as assessed by Euclidian or Mahalanobis [1936] distance metrics) to the features represented in *x*. The sample *x* is assigned the modal class of the nearest k neighbors.

For example, consider the data point represented by the question mark in Figure 4.3. Based on k = 5, it would be assigned the label associated with the most common class category of its five nearest neighbors (i.e., Class 1).



Gaussian kernels placed over each data point are used to estimate the distribution of features in each class.

Figure 4.3. K-nearest neighbor.

Small numbers of Gaussian kernels (dotted lines) are used to approximate the distribution of features in each class.

Parzen windows. Parzen windows (Parzen, 1967) are a generalization of the K-nearest neighbor technique. Instead of choosing the nearest neighbors and assigning a sample *x* with the label associated with the modal class of its neighbors, each vote is weighed by using a kernel function. With Gaussian kernels, the weight decreases exponentially with the square of the distance. As a consequence, faraway points become insignificant.

Kernel volumes constrain the region within which neighbors are considered. Consequently, Parzen windows are a better choice when there are large differences in the variability associated with each class. The data point (?) shown in Figure 4.4 is assigned to the dominant class in its immediate vicinity (i.e., class category 2).

A given feature vector is assigned the class label associated with the modal class of the k samples that are the most similar to it.

Figure 4.4. Parzen windows.

Gaussian mixture models (GMM). Gaussian mixture models provide a way to model the probability density functions of spectral features associated with each cognitive state. This can be accomplished using a superposition of Gaussian kernels (see Figure 4.5). The unknown probability density associated with each class or cognitive state can be approximated by the weighted linear combination of Gaussian density components. Given an appropriate number of Gaussian components and appropriately chosen component parameters (mean and covariance matrix associated with each component), a Gaussian mixture model can model any probability density to an arbitrary degree of precision. For more details, see Dempster, Laird, and Rubin (1977).

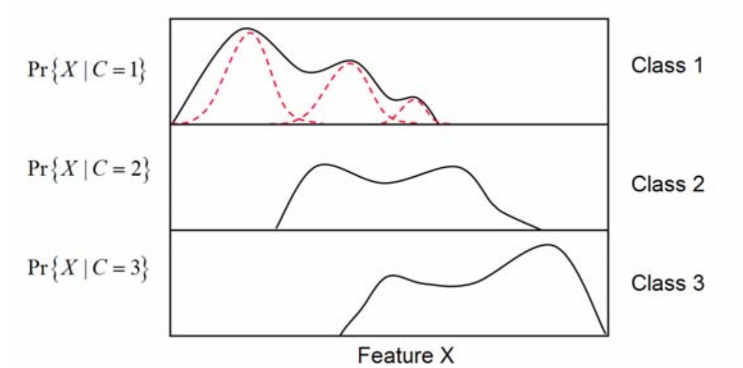


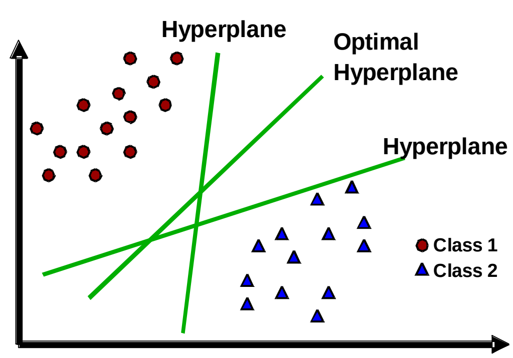
Figure 4.5. Gaussian mixture models.

These statistical classification techniques have an advantage over multilayer neural networks because they require minimal training time. KNN and Parzen windows require no training, whereas the GMM converges relatively quickly. KNN and Parzen window approaches require all patterns to be held in memory. Every new feature vector has to be compared with each of these patterns. However, despite the computational cost of these comparisons at run time, such systems have been shown to output classification decisions well within real-time constraints (Erdogmus, Adami, Pavel, Lan, et al. 2005).

Discriminative Classifier Approach

A discriminant function analyses approach has been employed in a fully operational mobile task evaluation, described later in the chapter. This approach used a support vector machine to discriminate between periods of low and high workload (Mathan, Whitlow, Dorneich, & Ververs, 2007).

Support vector machine (SVM). Support vector machines are linear classifiers that use a quadratic optimization procedure to find an optimal orientation and location for a discriminating hyperplane between classes. The optimization procedure finds a location and orientation for the hyperplane that lies as far as possible from examples in each class that are likely to be confused with each other (Figure 4.6).



Separating hyperplanes that are identified using the SVM procedure has been shown to maximize generalization performance.

Figure 4.6. Optimal hyperplane orientation could lead to better generalization. Adapted from Takahashi (2006).

Classification system uses a composite of three distinct classification approaches: K-nearest neighbor (KNN), Parzen windows, and Gaussian mixture models (GMM).

Separating hyperplanes that are identified using the SVM procedure has been shown to maximize generalization performance (Vapnick, 1999). Although they are linear classifiers, SVMs can be used to solve nonlinear problems by means of the so-called kernel trick. Data that may not be linearly separable in the original feature space can be projected into a high dimensional space where the data may be linearly separable (Figure 4.7).

Quadratic optimization procedures find an optimal orientation and location for a discriminating hyperplane between two classes.

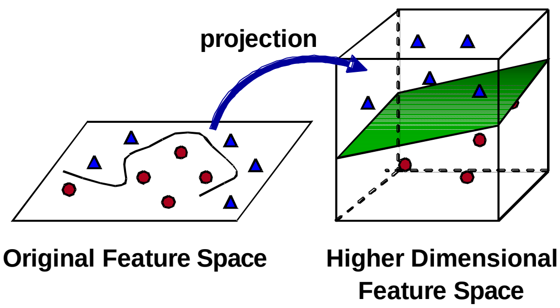


Figure 4.7. Transforms to higher dimensional space may result in separable data. Adapted from Takahashi (2006).

Fusion and Composite Techniques

Often it is possible to employ more than one sensor, or to employ more than one classification approach. There are two approaches, outlined next, that depend on whether you combine the input (i.e., sensor data) into the classifier or combine the output of the classifiers.

Sensor fusion. Sensor fusion uses multiple sources of sensor data to create the fusion at the sensor level before the discriminate features are calculated. This strategy for robust classification in noisy field environments integrates information from multiple sensor sources (assuming time synchronization) into a common feature vector that serves as input into a single classifier (see Chapter 7). Such an approach exploits the joint strengths of different data sources while minimizing their individual weaknesses.

Composite classifier fusion. Unlike sensor fusion (in which the fusion happens at the sensor input stage), composite classifiers fuse the output of multiple classifiers to create a final determination. A composite classification system (see Figure 4.8) has been developed that uses this technique. It employs three distinct classification approaches (K-nearest neighbor, Parzen windows, and Gaussian mixture models) and then fuses their outputs to make a final determination of cognitive state (Mathan et. al., 2005).

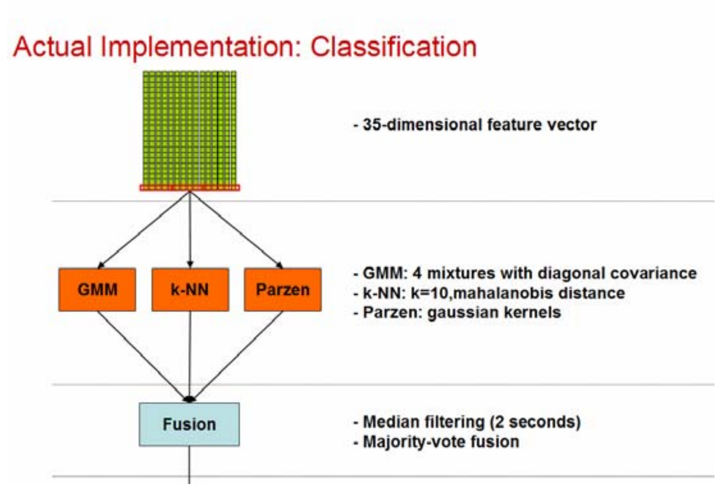


Figure 4.8. Composite classification system.

The composite classification system regards the output from each classifier as a vote for the likely cognitive state. The majority vote of the three component classifiers forms the output of the composite classifier. Fusing the outputs of multiple classifiers using a voting scheme is a widely used strategy to increase the robustness of a classification system. The equal weighting of different classifiers implicit in the voting scheme reflects the fact that no single classifier produces consistently superior results across participants and tasks.

Simple, vote-based fusion has been shown to improve the overall performance of classification systems (Kittler, Hatef, Duin, & Matas, 1998). There are a variety of alternative options for combining diverse classifiers. Exploring these options is an objective of future research.

Considerations in Evaluating Cognitive State Classification in Mobile Environments

Cognitive state classification can be achieved with a variety of methods. Each approach discussed earlier uses statistical pattern recognition techniques to define and, later, recognize unique classes of interest. The effectiveness of a particular cognitive state classification approach in mobile environments is framed by the following research issues:

- *Bias, variance, and temporal smoothing*
  - How well can the classifier fit and discriminate between workload classes in an inherently noisy and dynamic environment?
  - How well does the classifier generalize to unseen data over spans of tens of minutes—when task characteristics remain the same?
  - Can classification accuracy improve as the output of the classifier is integrated over time?
- *Discriminating features*: What aspects of signal serve to discriminate between high and low workload?

- *Fusion:* Can overall classification accuracy be improved by integrating additional sensor sources?
- *Sensor density*: How many channels (of EEG, for instance) are required for accurate classification?
- *Long-term generalization*: How well is the classifier likely to generalize over time spans of days as the task context and patterns of general physiological activity change (e.g., sleep, stimulants)?

Applied Exercise: Classification Design Considerations

Choice of classification approach is driven by multiple considerations. First and foremost, what is practical given the context? How many sensors need to be deployed? How much individualization is necessary, and how much training of classifiers will be required? Which classification approach is best able to discriminate between the cognitive states of interest? Within any one approach, how much tuning of parameters is needed to achieve good performance? Finally, how can the classification approach be meaningfully assessed to ensure that the resultant algorithms will allow for moment-to-moment classification? Design a classification approach that addresses these considerations.

Computational and Experimental Infrastructure

This section briefly outlines the computational and experimental infrastructure needed to classify cognitive state in an operational setting where mobility is a major challenge. In general, a system constructed for a mobile application environment to assess cognitive state classification algorithms consists of the following:

- *Sensors*: a variety of sensors to collect raw physiological and neurophysiological data.
- *Mobile processing*: mobile semirugged computer platforms to process the raw sensor data into cognitive state classification assessments.
- *Wireless data network*: a wireless data infrastructure to send the classification assessment to automation to close the loop, or to convey open-loop feedback of subordinates' state to human leaders.
- *Experimenter's base station*: a computing infrastructure and base station to control the IT component of the experiment and to troubleshoot any unexpected problems.

Design of Experiments

The evaluation of cognitive state classification algorithms in a mobile setting is fraught with several unique challenges not found in laboratory settings. First, it is much more difficult to design a task environment that reliably produces the cognitive state of interest when moving from a constrained task environment to a "free-play" operational environment. Second, because the task environment is not subject to the normal level of experimental control, it is much more difficult to know *ground truth* (i.e., the actual cognitive states experienced by the participant). Finally, the metrics used to assess the viability of the classification algorithms to distinguish the cognitive states of interest must take into account moment-to-moment discriminability, as opposed to averages of means over time (as discussed earlier).

Manipulating the Cognitive State of Interest

In addition to the practical and system configuration challenges faced when moving from the laboratory to field studies, there are issues of experimental control and the characterization of cognitive state in less constrained environments. It is essential to select tasks both that are operationally relevant and that afford reasonable adaptations that improve performance. In the laboratory it is possible to develop simple tasks in which the cognitive state of interest (e.g., cognitive workload) is manipulated precisely and consistently. Additionally, a user's performance can be collected and evaluated accurately. This makes it relatively easy to establish ground truth about a user's likely workload, for instance. However, when developing operationally relevant tasks in a field environment, it becomes substantially harder to manipulate workload precisely and to interpret and assess a user's performance without compromising operational realism.

In many operational settings, it is not always possible to vary workload directly. Instead, one must vary task load to induce cognitive workload. Furthermore, the amount of cognitive workload induced in a participant is a function of factors such as stress, fatigue, training, experience, and individual differences in capabilities. Thus, methods must be devised to correlate task load directly to workload in a systematic way in order to derive ground truth.

Ground Truth

In order to calculate the accuracy of a classification approach, classifier results are compared with ground truth. The output of the classifier at any moment is then compared with ground truth to determine the accuracy of the classifier.

The principal issue in scenario and task design is to create detectable and sustained periods (5–10 min) of high or low workload multiple times within any single data collection session. Definable periods of high and low workload sustained by participants are difficult to obtain directly from task characteristics, for the reasons discussed earlier. Thus, indirect methods must be used.

There are several classes of indirect methods: observation, secondary task performance, and participant self-reporting. Often, human experts can observe the experiment and determine ground truth based on their knowledge of task demands and the demonstrated behavior of the participant. When possible, secondary tasks can be introduced as discrete probes (in which metrics include performance and response latency) or as continuous tasks in which performance declines on the secondary tasks when the primary tasks induce higher workload. Participants can do a postscenario cognitive walkthrough, often with time-stamped videos, and report their self-assessment of their level of cognitive workload. These methods can be used individually or together to produce ground truth.

There are two important considerations when using some or all of these techniques. First, both the ground truth data and the classification data must be on a common (and accurate) time-stamp system, to allow for a moment-to-moment determination of classification accuracy. Second, for participants doing self-assessments, the terms *high workload* and *low workload* should be defined in an understandable manner.

- Operationally, *low workload* can be defined as times when the participant would have been able to take on additional cognitive tasks.
- *High workload* can be defined as times when it was not possible for the participant to take on any additional tasks and/or was not able to handle the current task load.

Classification Metrics

As discussed earlier, it is important to know how effectively a classification approach can differentiate between classes on a moment-to-moment basis. A metric used to evaluate classification performance is the Area under the Receiver Operating Characteristic (ROC) curve (see Duda, Hart, & Stork, 2001; see Chapter 6). ROC curves plot true positives (on the *y* axis) against false positives (on the *x* axis) as a threshold for discriminating between targets and distractors.

The ROC curve provides a way to assess the degree of overlap between two univariate distributions. It is widely used to evaluate human and machine signal detection capabilities. In addition, the ROC curve provides a way to assess the degree of overlap between the output of a classifier for two classes of data. Perfect classification produces an area under the curve value (Az) of 1.0, and chance performance produces an Az value of 0.5.

Test Your Knowledge

What is the motivation for using an ROC curve versus a simple statistical test for evaluating a significant difference between the means of distributions (classes)?

Tuning Classification Parameters

A major concern in the environments in which dismounted soldiers function is that noise from myriad sources could completely mask features that could be used to discriminate between high and low workload. Thus, a classifier may fail to discriminate adequately between workload classes. The capacity of a classifier to overfit training data is known as the *bias of the classifier*. These noise characteristics can also change dramatically over time—so that even if a classifier is able to effectively discriminate between workload classes over a short temporal window, it may fail to generalize adequately to unseen data collected a few seconds or minutes beyond the duration of the data used to train the classifier. The capacity of a classifier to generalize is referred to as the *variance of the classifier.*

One way to explore the bias and variance of a classifier is through a process called n-*fold cross validation*. This procedure entails splitting the data into *n* subsets. At each iteration of the validation procedure, one of these subsets (n_i) is used for testing the classifier, and the remaining $1 - 1/n$ sets are used for training the classifier. A typical choice of *n* is 10. Estimates of bias and variance get more conservative as the size of *n* decreases—the classifier has to be trained with less of the data and is assessed by generalizing to a larger subset of unseen data.

Advanced Warfighting Experiment

The general toolkit described in the previous section was realized in a mobile system that facilitated the evaluation of cognitive state algorithms (Dorneich, Mathan, Ververs, & Whitlow, 2007). The work discussed in the remainder of this chapter is grounded in an experiment conducted in an outdoor field environment. The Advanced Warfighting Experiment (AWE) was an evaluation of a MOUT (Mobile Operations in Urban Terrain) exercise at the U.S. Army Aberdeen Proving Grounds. The overall objective of the AWE was to evaluate the effectiveness of the toolkit's sensor-driven cognitive state assessment technologies in a realistic, operational, mobile environment.

Scenario Description

The AWE used a full Army platoon as participants. Of the 32 soldiers, four key leaders were instrumented with an augmented cognition system. There were two principal phases of the 12-day training session: part-mission training and full-mission execution. In part-mission training, the tasks changed each day, starting from simple entry techniques (e.g., door and wall breaching, upper-level entry, use of suppression devices), progressing to clearing techniques (e.g., room, hall, and stairwell entry and clearing; reflexive fire techniques), then on to defensive techniques (e.g., hasty defense of an urban area, security, protection, fields of fire), and finally to battle drills. Soldiers mastered a technique before moving to the next, as each technique built upon what was learned previously.

Battle drills were a culmination of all the training that soldiers received and enabled them to establish their own standard operating procedures. Examples included conducting a platoon attack, entering and clearing a building, reacting to an ambush, and securing at a halt. These tasks were not performed until the individual teams and squads demonstrated proficiency in all basic skills.

The second phase of the AWE was a full-mission evaluation involving a 24-hour training exercise. For this exercise, soldiers used techniques and skills learned during the part-mission training. The 24-hour period was divided into three 8-h phases:

1. Conduct dismounted movement along the lines of communication to an objective to ensure routes are free of mines and obstacles.
2. Conduct a cordon and search of the objective to kill, capture, or expel opposition forces operating in an urban area.
3. Prepare to defend the objective for an extended period and report any enemy activity in and around this key terrain.

This evaluation focused primarily on a platoon leader (PL), platoon sergeant (PSG), and two squad leaders (SL1 and SL2); however, the activities of their subordinates and responses from senior leaders had a direct impact on stress levels experienced by the leaders. The MOUT training facility, known as Mulberry Point, contained preassault staging and assault areas. It was a compound with several single- and multistory buildings with windows, doors, and hallways. This site served as a close combat training area.

The test site was equipped for data collection, including cameras in and around buildings. These data were used by experts in determining the workload ground truth. There were several stressors that the platoon-level training exercise introduced in the MOUT facility. These stressors were used to ensure that participants were placed in the cognitive states of interest (low and high workload). Each is summarized in Table 4.1.

In the remainder of this section, we address how the four principal challenges discussed in the "General Approach and Associated Toolkit" section (i.e., signal processing, classification, computational and experimental infrastructure, and design of experiments) were addressed in the AWE.

Table 4.1. Stressors Encountered by Soldiers in a MOUT Environment

Category	Example Stressors
Distributed operations	Distributed squads, loss of sight, reliance on faulty radios
Fatigue	Extended operational period (e.g., 24 h of operation): lengthy march followed by assault and then lengthy occupation of site in defensive posture
Realistic threats	Use of human OPFOR to prevent assault on urban facility, and to "hit" the friendly forces at different times; use of simunitions (soap bullets)
Evaluation stress	Evaluation of performance by commanders, Army trainers
Surprise / confusion	Unexpected elements imposed that affect plan, conditions, and mission; loss of communications and assets
Severe weather	Periods of high heat and humidity; intense rainfall
Information Gaps	Information flow variations from subordinates and commander

Signal Processing

Signal processing on the EEG signal was performed with a system that supported an independent signal-processing stream. Six channels were sampled at 256 samples/s with a bandpass from 0.5 Hz and 65 Hz (at 3 dB attenuation) obtained digitally with Sigma-Delta A/D converters. Quantification of the EEG in real time was achieved using signal analysis techniques that identified and decontaminated eyeblinks and identified and rejected data points contaminated with electromyographic (EMG), amplifier saturation, and/or excursions attributable to movement artifacts (see Berka et al., 2004, for a detailed description of the artifact decontamination procedures).

Decontaminated EEG was then segmented into overlapping 256 data-point windows called *overlays.* An epoch (the temporal window of analysis) consisted of three consecutive overlays. Fast-Fourier Transform (FFT) was applied to each overlay of the decontaminated EEG signal, multiplied by the Kaiser window ($\alpha = 6.0$), to compute the power spectral density (PSD). The PSDs were adjusted to take into account zero values inserted for artifact contaminated data points. The PSDs between 70 and 128 Hz were used to detect EMG artifact. Overlays with excessive EMG artifacts or with fewer than 128 data points were rejected. The remaining overlays were then averaged to derive a PSD for each epoch with a 50% overlapping window. Epochs with two or more overlays with EMG or missing data were classified as invalid. For each channel, PSD values were derived for each 1-Hz bin from 3 to 40 Hz and the total PSD from 3 to 40 Hz. Relative power variables were also computed for each channel and bin using the formula (total band power/total bin power).

Signal processing on the ECG signal focused on the importance of detecting the ECG peaks with a high degree of accuracy, accounting for missed peaks, and ignoring spurious peaks. Without appropriate correction, missing a single valid beat or adding a single spurious beat can lead to questionable estimates of spectral HRV measures. Heartbeat peaks within areas of the signal are identified with high-frequency content. False peaks are eliminated based on statistical comparisons to expected QRS morphology (QRS waves are related to the contraction of the left and right ventricles).

Scerbo et al. (2001) found that that the efficacy of HRV measures has mostly been limited to lab contexts, whereas IBI is a better measure in operational contexts. The IBI was computed as the time between successive peaks. The IBI was resampled at a rate of 256 Hz using a cubic spline interpolation. The resampled IBI was used as the workload indicator. Spectral methods, though not as effective, were also employed. The IBI signal was spectrally decomposed and the power in the band between .05 and .15 was used as an additional workload indicator.

Classification Approach

Cognitive state estimation was based on a support vector machine approach. The support vector machine used in this effort employed a radial basis function kernel with a kernel parameter of 1 and a slack parameter of .05.

The AWE evaluated the effectiveness of the classification algorithms to detect the user's cognitive state by correlating classification output to performance in various task load conditions. Experimentally, the principal hypothesis that was tested in the AWE was as follows:

> *The cognitive state classification algorithms would be able to differentiate periods of high and low cognitive workload using a combination of physiological (ECG) and neurophysiological (EEG) sensors.*

Computational and Experimental Infrastructure

Cognitive state classification was based on two sensor sources: EEG and ECG. An elaborate experimental infrastructure was developed to meet some of the challenges of collecting data in a harsh, mobile environment.

Sensors

EEG data were collected from the Advanced Brain Monitoring (ABM) EEG sensor headset (see Figures 4.9a and 4.9b). The sensor headset acquired six channels of EEG using a bipolar montage. Differential EEG were sampled from bipolar channels CzPOz, FzPOz, F3Cz, F3F4, FzC3, and C3C4 obtained digitally. Data were transmitted across a Bluetooth RF link to the collection laptop via an RS232 interface.

The Sensor Headset was developed by ABM as a portable system to record EEG signals. The headset fit snugly on the head and housed EEG sensors. It was important to minimize the movement of sensors to reduce signal artifacts. Snug-fitting caps reduced signal noise caused by sheering of the sensors against the head and scalp. Physiological recordings were made with an experimental eight-channel digital physiological recorder with low-powered EEG and EOG amplifiers designed specifically for ambulatory recordings. Amplification at the electrode site was important to boost the signal-to-noise ratio.

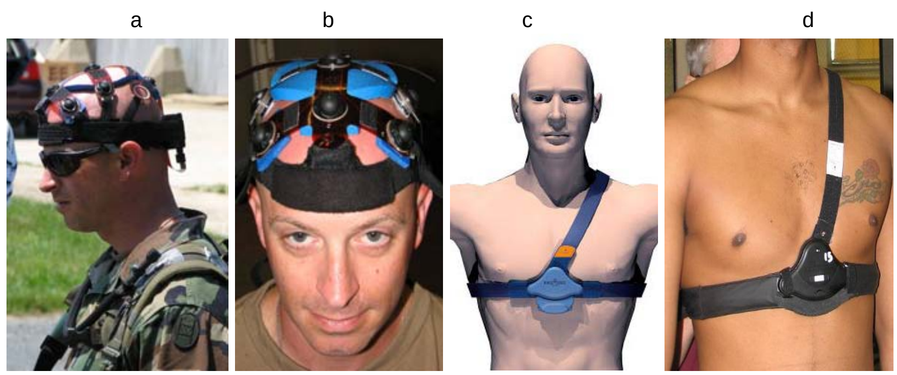


Figure 4.9. (a & b) ABM's wireless EEG Sensor Headset, (c & d) Hidalgo VSDS for ECG.

The Hidalgo Vital Signs Detection System (VSDS) measured heart rate, respiration rate, and body motion and position (see Figures 4.9c and 4.9d). Both waveform and summary data were transmitted across a Bluetooth communications link. The AWE used the ECG waveform (two views, sampled at 256 Hz) and three-axis accelerometry waveform (sampled at 25.6 Hz) signals.

Mobile Processing and Data Collection Platform

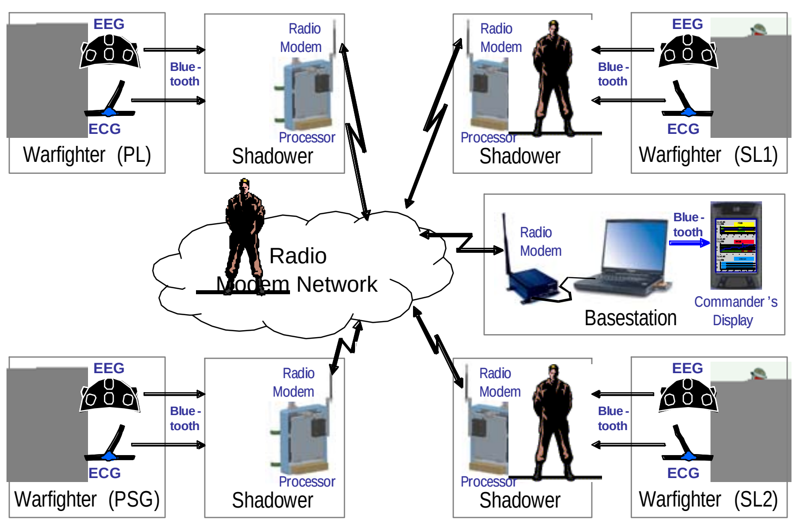
Each of the four primary soldier participants (PL, PSG, SL1, and SL2) was followed by a member of the experimental personnel in the role of shadower. Each shadower remained within a 30-m range of his/her participant to ensure Bluetooth connectivity. Each shadower carried a specially designed backpack (based on the MOdular Lightweight Load-carrying Equipment [MOLLE] system), which contained a Panasonic Toughbook® CF-51 computer equipped to receive Bluetooth communication from the participant's EEG and ECG sensors and audio from a wireless microphone. In addition to logging data, shadowers processed raw sensor data on their computers using Honeywell's Cognitive State Classification algorithms to produce a real-time assessment of their participants' cognitive state. That cognitive state assessment was then transmitted to the base station via a wireless data network (see next section).

The use of shadowers allowed sufficient computing power to process the raw data from the sensors without interfering with the mobility of the participant, as well as protecting the equipment from being broken during the more physical aspects of the missions. Ideally, small mobile, ruggedized systems configured with sufficient computing power worn on the body by the soldier would be used once parameters are downselected through the research process. Additionally, the shadower wore a webcam and logged video to the computer for later review by experts to enable determination of the workload ground truth. The participant wore a wireless microphone, and the resultant audio stream was multiplexed into the webcam video.

The base station received data from the four shadowers' computers via the wireless data network. The base station was the test team's command and control center for the devices and facilitated the diagnosis of problems, resetting of systems, and monitoring of system status. The base station performed several functions, including remote control of the four shadower computers (ability to stop/start processes), monitoring of processes on four shadower computers, running of the master radio, remote troubleshooting of the shadower computers, data collection, and the shutting down of processes at the end of a trial.

Wireless Network

The AWE employed a 900-MHz radio modem system to create a wireless data network connecting the four shadowers' computers to the base station. The ABM EEG and the Hidalgo VSDS communicated to the shadower computer via Bluetooth. Figure 4.10 illustrates the final data collection system and experimental infrastructure configuration.



Creating a stable and robust experimental infrastructure is key to successfully field-testing classification approaches.

Figure 4.10. Final data collection system and experimental infrastructure.

Several practical challenges were encountered during the AWE. First and foremost, the pace of the training was subject to each soldier's progress through a predefined set of tasks, drills, and procedures. Soldiers were trained to performance on battle drills. The use of "simunitions" (soap bullets) implied that all hardware—including potentially sensitive equipment such as EEG sensors—had to be hardened to withstand a direct hit of a simunition round. During the experiment, the ABM EEG system sustained a direct hit but was not damaged.

The weather was another challenge: During two of the days of training, 12 inches of rain fell. The AWE required that wireless connectivity be maintained over two networks, the Bluetooth connections (between sensors and shadowers) and the 900-MHz radio modem network. Power consumption of the mobile equipment was always a challenge, and battery management was key to ensuring that all devices continued to function despite inevitable delays and schedule changes.

Finally, EEG sensor integration with soldier's standard equipment was a challenge that required special modifications to the padding and padding configuration under the soldiers' helmets.

Design of Experiments

The independent variable in the AWE was workload (for all phases). The experimental scenarios were manipulated to ensure definable periods of high and low cognitive workload. Low-workload periods were characterized by engagement in a single task that was well within the current cognitive capability of the soldier, usually under little or no stress or time pressure. Periods of low workload included completing initial paperwork, reporting activities, preplanning, establishing a hasty defense position (e.g., foxhole), consolidation/transition, after-action reviews, and periods of low activity during missions.

High-workload periods were characterized by multiple-task performance, often under time pressure and fatigue. Examples of high workload included replanning caused by changing circumstance (e.g., enemy location, available squads, loss of communication), directing squad movements during preassault, assault, managing multiple communications (i.e., responding to commanders, squad leaders, or other platoon leaders), or calling for fire. Stressors that contributed to high workload included frustration, loss of communication, lack of asset availability, and loss of situation awareness of squad locations and activities.

Ground Truth

During the AWE, multiple streams of data were collected with the objective of providing experts with enough insight to make a determination of ground truth levels of workload for each participant in each scenario. Data included video from a roaming camcorder (focused on the platoon-level action), video from the webcam of the shadower (focused on the participant), notes from an observer at a central (video) monitoring site, annotations radioed in from the shadower and entered at the base station into the time-stamped data stream via an annotator's interface, postscenario cognitive walthroughs with the participants as they reviewed (with an experimenter) the video of the day's events, postscenario NASA TLX surveys (Hart & Staveland, 1988), and questionnaires.

Not all data were collected for every part-mission and full-mission scenario, but some combination of data streams was available for expert review. The notes, annotations, and cognitive walkthrough feedback data streams were merged (by time stamp) into a spreadsheet. An expert then reviewed the video streams, taking into account the various data sources, to make a moment-to-moment assessment of the cognitive workload being experienced by the participant at any given time stamp. The result was a time-stamped series of blocks of low, medium, or high cognitive workload. Physical load was also assessed by the experts.

Two experts independently performed the ground truth analysis described earlier. Their respective results were then compared to gain a measure of interrater reliability on the cognitive workload assessments of ground truth. For the data sets analyzed, agreement between the raters was high: Agreement in the rating of physical load was 94.9%, and agreement in the rating of cognitive workload was 87.9%.

A final, canonical, assessment of ground truth was created by reconciling the two individual expert's assessments. Periods of disagreement were flagged. The two experts then jointly reviewed the video and other data streams to make a final assessment of the workload in the disputed block. In cases in which no consensus was reached, a third rater was available to resolve the disagreement; however, this option was never needed. Reconciled ground truth tables were used to calculate the accuracy metric of the classification algorithms.

Test Your Knowledge

The AWE was an evaluation in a mobile, operational context using participants (i.e., soldiers) performing their natural, domain-specific tasks. The elaborate setup described was designed to meet four challenges, which were made particularly difficult in a mobile setting (i.e., signal processing, classification, computational and experimental infrastructure, and design of experiments). The objective of the evaluation was to assess workload classification techniques during multiple operational tasks requiring different levels of cognitive and physical engagement. How were these challenges overcome? Which strategies were the most effective in overcoming these challenges?

Lessons Learned

The AWE data analysis forms a good example of the specific lessons learned when evaluating the general approach outlined earlier in the chapter. For a more extensive description of classification results, see Mathan, Whitlow, Dorneich, and Ververs (2007). The lessons derived from the results reviewed here fall into three principal categories: (a) choosing parameters of the classification approaches to improve performance in mobile environments, (b) creating a set of features that adequately captures the cognitive state of interest, and (c) improving classification while minimizing computational demands.

Tuning Classification Parameters

In noisy operational environments, EEG and other electrophysiological sensors could be compromised by noise over short temporal windows. One strategy for dealing with momentary fluctuations in classification accuracy is to median-filter the output of the classifier over different time windows. One consequence of such temporal smoothing of classifier output is it may introduce a lag in the decision process. The analysis must consider the trade-off in accuracy as the temporal window of output smoothing is varied.

The classification approach was assessed with two individuals, the platoon leader and the platoon sergeant, both employing the widely used tenfold cross-validation approach and the more conservative twofold cross-validation procedure.

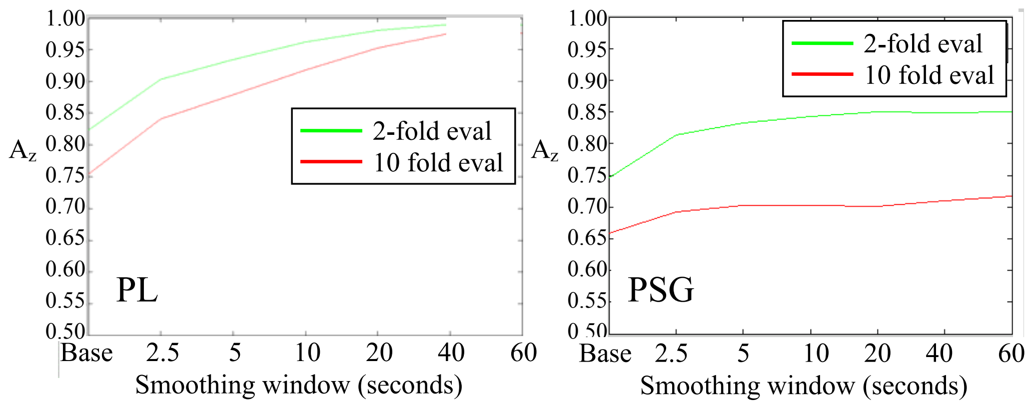


Figure 4.11. EEG-based classification accuracy for the PL (left) and PSG (right) as a function of validation technique and temporal smoothing window.

As Figure 4.11 illustrates, base EEG classification accuracy for the PL ranged from 0.76 (using twofold cross validation) to 0.83 (using tenfold cross validation). Base results for the PSG ranged from 0.66 (using twofold cross validation) to 0.75 (using tenfold cross validation), as seen in Figure 4.11 (right). Accuracy for both soldiers rose monotonically up to a 1-min-long temporal smoothing window. However, the rate at which temporal smoothing benefited accuracy diminished beyond approximately 2 to 3 s of smoothing. This analysis confirms the lesson learned that the single-point analysis of classifier accuracy does not convey the bias, variance, and generalizability of a classifier approach.

The discrepancy between the more conservative twofold validation and more optimistic tenfold cross validation was more pronounced for the PSG than it was for the PL. This could indicate some change in the features that serve to discriminate between high and low workload over time; these changes could stem from changes in task, strategy, artifacts, or a variety of physiological factors.

Discriminating Features

The analysis included a qualitative examination of the spectral features that serve to discriminate between high and low workload. Figure 4.12 depicts the power spectral densities (PSD) for high and low workload across six channels of EEG for the PL (left) and the PSG (right).

Each graph in Figure 4.12 represents a channel. The *x* axis in each graph represents frequency, and the *y* axis represents amplitude. One line in each graph represents average spectral power in the high-workload condition; another line represents average spectral power in the low-workload condition. Finally, the center line in each graph corresponds to the mean spectral power across both high- and low-workload conditions. Qualitatively, the key distinction is the separation (if any) between the high and low spectral power lines.

An analysis of the graphs for both participants suggests that power in the beta (12 to 30 Hz) and gamma (30 to 40 Hz) bands is the most discriminative feature for both participants. However, this pattern is most pronounced for PL and may account for the superior classification results observed relative to PSG. This discrepancy across individuals also points to the lesson of the importance of an individualized approach to classification, rather than an approach that relies on group norms.

Review: Techniques for Improving Classification in Mobile Environments

Mobile environments require that the number of sensors and the computational demands of the classification algorithms are minimized while accurate classification is maintained in the presence of motion-induced noise artifacts. Robust classification is addressed by sensor fusion techniques. Minimizing computational demand requires that the minimum number of sensors be identified.

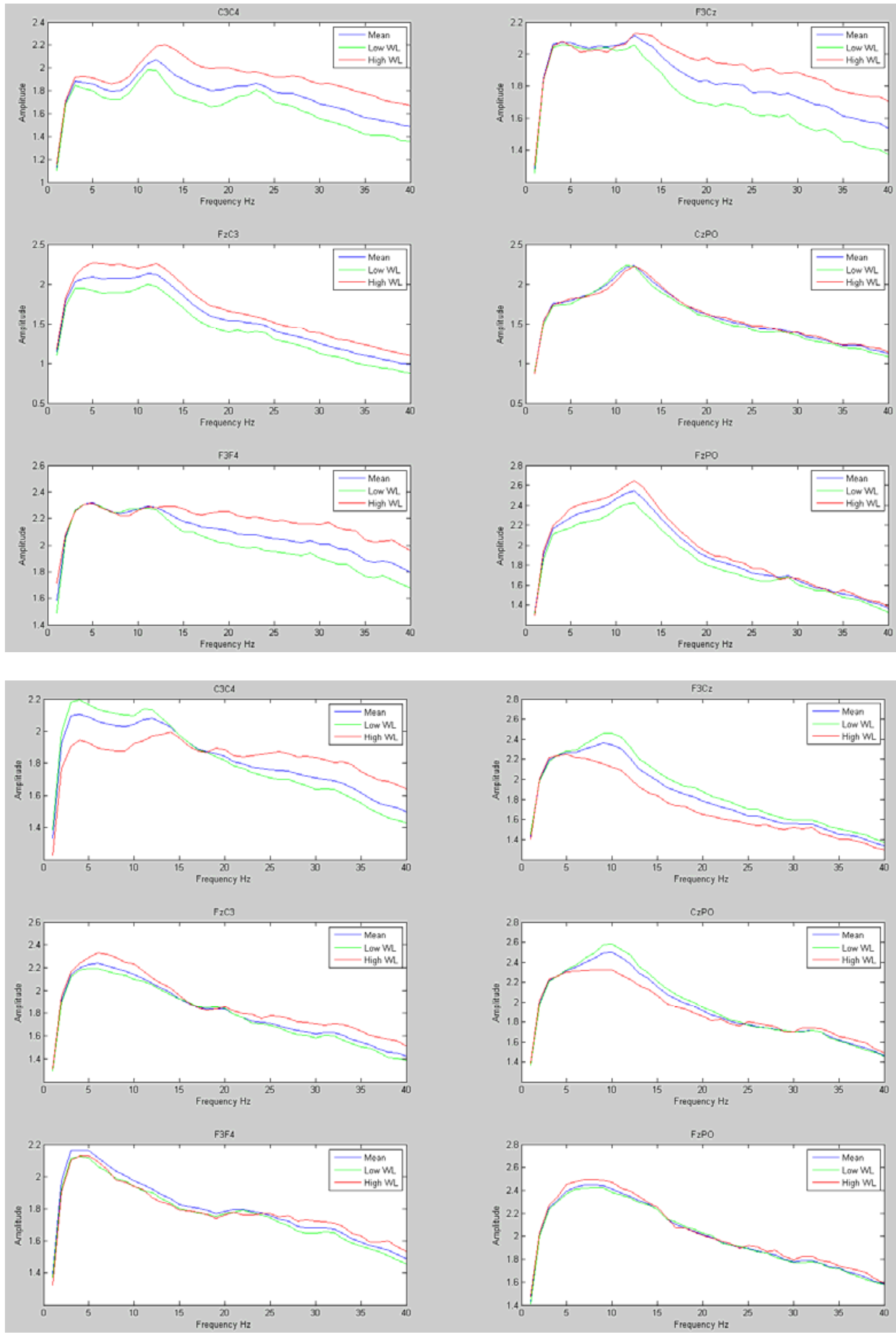


Figure 4.12. Power spectral densities in each band for the PL (left) and PSG (right).

Sensor Fusion

One lesson learned in the work reported here was the utility of fusing data from multiple sensor sources to improve classification in noisy field environments. Such an approach exploits the joint strengths of different data sources while minimizing their individual weaknesses. Fusing multiple sensor sources into a common feature vector allows a classifier to find an optimal weighting for each feature based on training data.

We assessed the effect of including IBI estimates as a feature for classification. The fusion of cardiac data provided a substantial boost to overall classification performance—these improvements were most pronounced for PSG, as seen in Figure 4.13. Base classification for PL went up from 0.76–0.83 to 0.87–0.95, and base classification for PSG went up from 0.66–0.75 to 0.83–0.86.

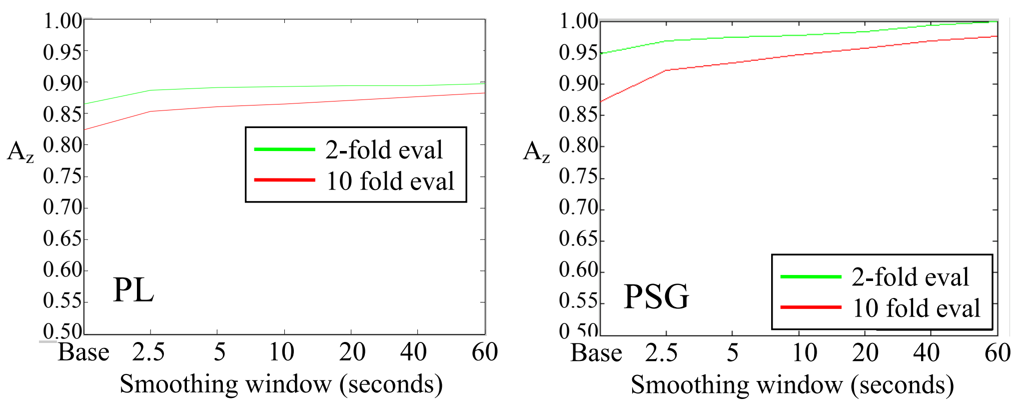


Figure 4.13. Classification accuracy for the fused sensor data for the PL (left) and PSG (right).

Sensor Density

The EEG system used in the field evaluation consisted of a six-channel system. In a mobile setting, such as dismounted soldier operations, it is important to reduce the number of sensors to the minimum required to capture the underlying cognitive state of interest. An analysis was conducted to identify a subset of the six EEG channels that could match or exceed the performance of all channels together. With each iteration of a backward elimination-ranking algorithm, each channel of the current set was sequentially eliminated from consideration. The channel whose exclusion led to the best performance results was eliminated from further consideration.

The ranking assigned to each channel corresponded to the order in which it was eliminated. The first channel to be eliminated was ranked as being last in importance, whereas the last channel to remain was regarded as being of the highest importance. Performance of each feature subset was assessed using tenfold cross validation. The performance metric used was the area under the Receiver Operating Curve (Az; see Chapter 6). The channel-ranking procedure produced the channel ranks shown in Figure 4.14 (PL left and PSG right), which plots classification accuracy as a function of the top *n* channels.

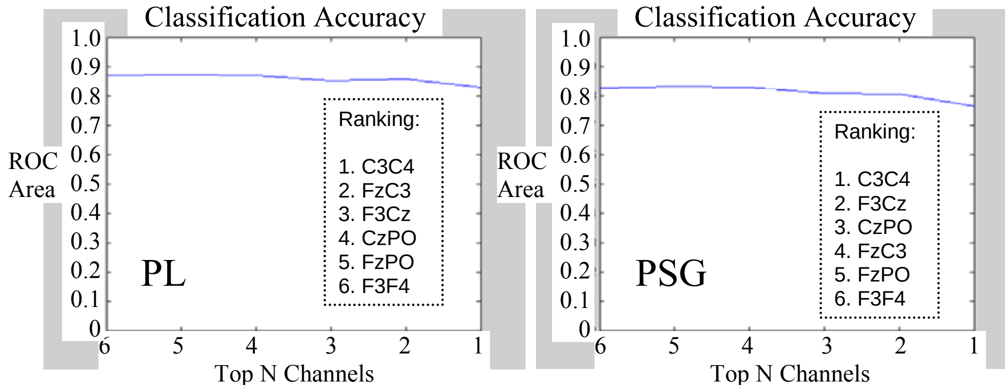


Figure 4.14. Classification accuracy as a function of the top *n* channels.

The channel-ranking procedure yielded a consistent set of features for both participants. Classification performance suffered little with the exclusion of all but the two most salient channels. The top channels were identical for both participants (C3C4). These channels, which were located at the apex of the skull, are likely to have been least affected by helmet-related artifacts because of good clearance between the sensors and the helmet at this location. The lesson learned was the importance of stable electrode sensor sites, which is possible even in noisy conditions.

Although these results require further validation, the lesson learned was that accurate workload classification may be feasible with as few as one or two sensors. This has compelling implications for the design of practical EEG systems that could be integrated easily within helmets and find broad user acceptance.

Best Practices

The best practices derived from this work stem from the approaches taken to overcome four principal challenges: (a) collecting and processing EEG signals in a mobile context; (b) developing classification algorithms that address individualization, bias, and generalizability; (c) designing experiments to assess classification in a domain-relevant operational setting with soldiers; and (d) building a computational and experimental infrastructure to support assessment. For each challenge, a table of best practices and practical recommendations is given.

Meeting the Mobility Challenge

Table 4.2 outlines some of the best practices, guidelines, and recommendations in several areas when meeting the challenge of collecting and processing EEG signals in a mobile context.

Table 4.2. Best Practices, Guidelines, and Recommendations Addressing Mobility Challenges

Area	Best Practices	Guidelines / Recommendations
Equipment	Develop the capability to collect data in the actual environment.	Select an EEG system that preamplifies the signal at the electrode site to enable low noise

		measurements. Cabling between the sensor and data collection equipment must be secured to avoid cable sway-induced noise artifacts.
Signal processing	Understand the noise artifacts, and understand the signal of interest.	Understand the noise artifacts by running a series of experiments in which you control (in turn) every type of mobility condition (e.g., stationary, standing, walking, running, head motion). Understand the signal of interest by first collecting data in pristine environments in order to be able to identify the signal later amid all the noise artifacts introduced by the operational environment.
Signal processing	Develop stability controls to improve adaptive filtering (see Chapter 7). When faced with the extreme artifacts in a mobile environment, most adaptive filters would become unstable and unusable.	If the benefits of adaptive filter algorithms are to be obtained in a mobile environment, the algorithms must be stabilized during high-amplitude spikes. See, for instance, Mathan, Dorneich, and Whitlow (2007).
Signal processing	Findings from prior research were quickly identified as inadequate for identifying relevant EEG sites for use in applied operational domains.	Run pilot studies in the operational environment that use the same or a similar task to identify the cognitive states of interest. Start with many EEG sites and run sensor density analysis to rank the channel contributions.
Signal processing	Collect sufficient data to determine how much training data are required to provide good classification performance.	Use pilot studies to determine how much training data are needed. The amount of data needed varies depending on the nature of the task environment, signal-to-noise ratio, and classification techniques used.

Meeting the Classification Challenge

Table 4.3 outlines some of the best practices, guidelines, and recommendations in several areas when meeting the challenge of developing classification algorithms that address individualization, bias, and generalizability.

Table 4.3. Best Practices, Guidelines, and Recommendations Addressing Classification Challenges

Area	Best Practices	Guidelines / Recommendations
Classification	Fit the approach to the constraints of the environment.	Determine the spatial density of EEG sensor arrays based on an understanding of the nature of the tasks, pace of task switching, and specific types of cognitive processing involved. Consider the constraints imposed by sensor density, computational efficiency, precise task adaptation needs, and the desire for a high degree of classification accuracy during ongoing research studies.
Classification	Explore multiple temporal windows.	Temporal smoothing should be employed to stabilize classifier output. Classifier update rates need only satisfy the requirements of the pace of adaptation switching.
Sensor density	Determine the ideal number of	Once the classifier approach goes beyond the

	sensors by considering processing demands, the operational environment, and the generalizability of classification across multiple situations.	most informative features (site by frequency band) the classifier begins to overfit to noise and degrade classification performance, much as does adding unnecessary parameters to a regression model.
Sensor location	Choose the sensor location based on the cognitive state of interest.	Generic workload assessment can employ a low-density array over the frontal central and parietal lobes. Data from sensors located in the occipital area tend to be noisier and do not provide discretionary information, given that most tasks involve visual processing.
Sensor type	Choose the sensor type based on the cognitive state of interest.	If you need immediate feedback on specific events, use a time-locked EEG measure such as evoked response potentials (ERPs). If you need general task loading over extended periods, use oscillatory EEG measures such as PSDs.
Fusion	Utilize complementary measures of cognitive state where appropriate.	ECG provides information on tonic states (i.e., slowly changing), whereas EEG provides high temporal fidelity (i.e., moment-to-moment). Together, the two have been shown to improve classification accuracy.

Meeting the Infrastructure Challenge

Table 4.4 outlines some of the best practices, guidelines, and recommendations in several areas when meeting the challenge of building a computational and experimental infrastructure to support assessment.

Table 4.4. Best Practices, Guidelines, and Recommendations Addressing Infrastructure Challenges

Area	Best Practices	Guidelines / Recommendations
System integration	Ruggedize the equipment for testing in a field environment.	Use ruggedized laptops that come with shock-mounted hard drives to protect your data, and include effective thermal management.
System integration	Ruggedize all connections.	Secure all cable connections. For instance, typical USB connectors were not designed to maintain a connection under mobile conditions.

Meeting the Assessment Challenge

Table 4.5 outlines some of the best practices, guidelines, and recommendations in several areas when meeting the challenge of designing experiments to assess classification in a domain-relevant operational setting with soldiers.

Table 4.5. Best Practices, Guidelines and Recommendations Addressing Assessment Challenges

Area	Best Practices	Guidelines / Recommendations
Task definition	Consult domain experts.	Even if it is not possible to perform the actual task in early experiments, developing representative tasks lends confidence that the findings will be transferable to the actual domain. Not only does designing tasks with input from domain experts

		save considerable time, but results will be better received because of their ecological validity.
Task definition	Baseline tasks early and often to ensure that representative participants perform and perceive different task loads as low and high.	Task load does not produce the same workload in different participants, or even in the same participant over time. Maximize discrepant task loading for good binary classification.
Experimental control	Evaluation of techniques in operational environments often results in a loss of experimental control as evaluations move from the lab to the field.	Free-play evaluations with high ecological validity are very effective in loading leaders with varying levels of workload. Even in a free-play evaluation, an operator or controller can manipulate the workload of participants by changing the scenario, introducing unexpected events, and controlling the pace of operations.
Experimental design	Whenever possible, simplify the experimental design to reduce the complexity of conducting field studies.	Inevitably, the system integration phase will take three times longer than expected. Limit the number of research questions of interest and avoid rolling up everything into a single study.
Risk management	Consider an experimental design that includes segments with severable benefits.	Ensure that data analysis is possible on the cumulative data collected (i.e., each day's data), so if data collection becomes impossible, the experiment can still produce results on whatever data were collected thus far.
Ground truth	Explicitly design the data collection plan for ground truth.	If you are videoing the participant, make sure that a microphone channel is included, as it is often difficult to decipher the state of the participant from video alone. Reviewing video (and audio) with the participant immediately after the experimental trial provides the single best data source for insight into the participant's cognitive loading at any given moment. Make sure that all data streams share a common time stamp.

Design Guidelines

Design guidelines take the form of overarching considerations that become important when cognitive state classification work is matured outside the laboratory and is used in real-world, mobile, operational contexts. Table 4.6 captures a principal guideline in each of the four challenge areas.

Table 4.6. Design Guidelines

Area	Design Guideline
Mobility	Thorough advanced signal-processing algorithms are essential to ensure a clean signal for cognitive state classifiers. It is particularly important to remove or identify noise artifacts in harsh operational environments.
Classification	There is not a one-size-fits-all approach to cognitive state classification. Individualized measurements are necessary for each individual. In addition, because of changes in physiological data over time, regularly scheduled baselines will need to be captured to maintain a high level of classification accuracy.

Infrastructure	There is a need to further ruggedize physiological and neurophysiological sensors and sensor systems to enable the deployment of this capability.
Assessment	The assessment of classification effectiveness will always require an evaluation to capture the context of the mission and task and incorporate user feedback as a basis of ground truth information. In addition to a complete understanding of the target environment, thorough interviews with participants and multiple raters of ground truth classification will help to minimize errors in cognitive state classification caused by poor insight into the cognitive loading requirements of the task environment.

Parting Message

The evaluation of cognitive state classification techniques outside the laboratory is wrought with challenges apart form the classification techniques themselves. Successful assessment will depend on the ability to collect valid signals robustly in a noisy environment, and require a computing and experimental infrastructure that can enable realistic experiments in the domain of use. The design of these experiments is itself a major challenge, as one no longer has the benefit of well-defined, well-understood laboratory tasks that engage the cognitive state of interest. Failure to address any of these challenges severely compromises the ability to draw meaningful conclusions about the use of cognitive state classification algorithms in the target operational domain.

References

Aasman, J., Mulder, G., & Mulder, L. J. M. (1987). Operator effort and the measurement of heart rate variability. *Human Factors, 29*, 161–170.

Backs, R. W., & Seljos, K. A. (1994). Metabolic and cardiorespiratory measures of mental effort: The effects of level of difficulty in a working-memory task. *International Journal of Psychophysiology, 16(1)*, 57–68.

Balles, J. (1988, December). Vincennes: Findings could have helped avert tragedy: Scientists tell Hill panel. *APA Monitor,* pp. 10-11.

Banks, S., & Lizza, C. (1991, June). Pilot's associate: A cooperative, knowledge-based system application. *IEEE Expert,* pp. 18–29.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*(2), 276–292.

Beh, H. C. (1990). Achievement motivation, performance and cardiovascular activity. *International Journal of Psychophysiology, 10(1)*, 39–45.

Berka, C., Levendowski, D., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., Zivkovic, V. T., Popovic, M. V., & Olmstead, R. (2004). Real-time analysis of EEG indices of alertness, cognition, and memory acquired with a wireless EEG headset. *International Journal of Human-Computer Interaction, 17*(2), 151–170.

Boiten, F. A. (1998). The effects of emotional behavior on components of the respiratory cycle. *Biological Psychology, 49*(1–2), 29–51.

Buller, M. J., Hoyt, R. W., Ames, J., Latzka, W., & Freund, B. (2005). Enhancing warfighter readiness through physiologic situational awareness—The warfighter physiological status monitoring—Initial capability. *Proceedings of the 1st International Conference on Augmented Cognition. pp. 335-344*. Mahwah, NJ: Erlbaum.

Cochran, E., Miller, C., & Bullemer, P. (1996). Abnormal situation management in petrochemical plants: Can a pilot's associate crack crude? In *Proceedings of the NAECON, 1996* (pp. 806–813), Houston, TX**.**

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, 39*, 1–38.

Dorneich, M. C., Mathan, S., Ververs, P. M., & Whitlow, S. D. (2007). An evaluation of real-time cognitive state classification in a harsh operational environment. In *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting* (pp. XXX–XXX). Santa Monica, CA: Human Factors and Ergonomics Society.

Dorneich, M. C., Ververs, P. M., Mathan, S., & Whitlow, S. D. (2005, October). A joint human-automation cognitive system to support rapid decision-making in hostile environments. Presented at the International Conference on Systems, Man and Cybernetics, Waikoloa, HI.

Duda, R. O., Hart, R. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). New York: Wiley.

Erdogmus, D., Adami, A., Pavel, M., Lan, T., Mathan, S., Whitlow, S., & Dorneich, M. (2005, March). *Cognitive state estimation based on EEG for augmented cognition.* Presented at the 2nd IEEE EMBS International Conference on Neural Engineering, Arlington VA.

Future Force Warrior. (2004). Retrieved August 11, 2008, from http://nsrdec.natick.army.mil/about/techprog/index.htm

Garavan, H., Ross, T. J., Li, S.-J., & Stein, E. A. (2000). A parametric manipulation of central executive functioning using fMRI. *Cerebral Cortex, 10*, 585–592.

Geddes, N. D. (1985, September). Intent inferencing using scripts and plans. In *Proceedings of the First Annual Aerospace Application of Artificial Intelligence Conference,* Dayton, OH, pp. 160-172.

Gevins, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex, 10*, 829–839.

Gevins, A., & Smith, M. (2003). Neurophysiological measure of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science, 4*(1–2), 113–132.

Girolamo. (2005). Augmented cognition for warfighters: A beta test for future applications. *Proceedings of the 1st International Conference on Augmented Cognition*, Mahwah, NJ: Lawrence Erlbaum Associates.

Hart, S. G., & Staveland, L. E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human mental workload*. pp. 138-183, Netherlands: Elsevier.

Heslegrave, R. J., & Furedy, J. J. (1979). Sensitivities of HR and T-wave amplitude for detecting cognitive and anticipatory stress. *Physiology & Behavior, 22*(1), 17–23.

Hockey, G. R. J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. II). New York: Wiley, pp. 44/1-44.

Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., & Chance, B. (2004). Functional optical brain imaging using near-infrared during cognitive tasks. *International Journal of Human-Computer Interaction, 17*(2), 211–227.

Kalsbeek, J. W. H., & Ettema, J. H. (1963). Scored irregularity of the heart pattern and measurement of perceptual or mental load. *Ergonomics*, *6*, 306–307.

Kittler, J., Hatef, M., Duin, R.P.W & Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(3), 226–239.

Kramer, A. (1991). Physiological metrics of mental workload: A review of recent progress. In D. Damos (Ed.), *Multiple task performance* (pp. 279–328). London: Taylor & Francis.

Krebs, H., Brashers-Krug, T., Rauch, S., Savage, C. R., Hogan, N., Rubin, R. H., Fischman, A. J., & Alpert, N. M. (1998). Robot-aided functional imaging: Application to a motor learning study. *Human Brain Mapping*, *6*, 59–72.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science of India, 12,* 49–55.

Makeig, S., & Jung, T -P. (1995). Changes in alertness are a principal component of variance in the EEG spectrum. *NeuroReport, 7*(1), 213–216.

Mathan, S., & Dorneich, M. (2005, July). *Augmented tutoring: Improving military training through model tracing and real-time neurophysiological sensing*. Presented at the 11th International Conference on Human-Computer Interaction (Augmented Cognition International), Las Vegas, NV.

Mathan, S., Dorneich, M. C., & Whitlow, S. D. (2007). *Statistical control of adaptive ocular filter stability* (U.S. patent application #20070239814). Washington, DC: U.S. Patent and Trademark Office.

Mathan, S., Mazaeva, N., Whitlow, S., Adami, A., Erdogmus, D., Lan, T., & Pavel M. (2005, July). Sensor-based cognitive state assessment in a mobile environment. Presented at the 11th International Conference on Human-Computer Interaction (Augmented Cognition International), Las Vegas, NV.

Mathan, S., Whitlow, S., Dorneich, M., & Ververs, T. (2007). Neurophysiological estimation of interruptibility: Demonstrating feasibility in a field context. *Proceedings of the 4th International Conference of the Augmented Cognition Society*. Baltimore MD.

Mikulka, P., Hadley, G., Freeman, F., & Scerbo, M. (1999). The effects of a biocybernetic system on vigilance decrement. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. XXX–XXX]. Santa Monica, CA: Human Factors and Ergonomics Society.

Miller, C. A., & Dorneich, M. C. (2006). From associate systems to augmented cognition: 25 years of user adaptation in high criticality systems. In *Proceedings of the 2nd Augmented Cognition International*. San Francisco, CA.

Miller, C., Wu, P., Kirchbaum, K., & Kiff, L. (2004). Automated elder home care: Long term adaptive aiding and support we can live with. In *Proceedings of the AAAI Spring Symposium on Interaction between Humans and Autonomous Systems over Extended Operation*. 22-24 March. Stanford, Palo Alto, CA., U.S.A.

Montain, S. J., Sawka, M. N., & Wenger, C. B. (2001). Hyponatremia associated with exercise: Risk and pathogenesis. *Exercise Sports Science Review*, *29*, 113–117.

Neumann, D. L. (2002). Effect of varying levels of mental workload on startle eyeblink modulation. *Ergonomics, 45*, 583–602.

Parasuraman, R. (2003). Neuroergonomics: Research and practice. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 5–20.

Parasuraman, R., Mouloua, M., & Hilburn, B. (1999). Adaptive aiding and adaptive task allocation enhance human-machine interaction. In M. W. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 129–133). Mahwah, NJ: Erlbaum.

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies, 59*(1–2), 185–198.

Parzen, E. (1967). On estimation of a probability density function and mode. *Time Series Analysis Papers*. San Diego, CA: Holden-Day.

Pope, A. T., Bogart, E. H., & Bartolome, D. (1995). Biocybernetic system evaluates indices of operator engagement. *Biological Psychology, 40*, 187–196.

Porges, S. W., & Byrne, E. A. (1992). Research methods for measurement of heart-rate and respiration. *Biological Psychology*, *34*(2–3), 93–130.

Porges, S. W., & Raskin, D. C. (1969). Respiratory and heart rate components of attention. *Journal of Experimental Psychology, 81*, 497–503.

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2000). A closed-loop system for examining psychophysiological measures for adaptive automation. *International Journal of Aviation Psychology*, *10*, 393–410.

Robertson, G. (2000). Flight demonstration of an associate system—A rotorcraft pilot's associate example. In *Proceedings of the AHS International, Annual Forum*. May 2-4, Virginia Beach, VA, USA., Vol, 56, pp. 431–445.

Scerbo, M. W., Freeman, F. G., Mikulka, P. J., Parasuraman, R., DiNocero, F., & Prinzell, L. J. III. (2001). The efficacy of psychophysiological measures for implementing adaptive technology. *NASA Tech. Report NASA/TP-2001-211018*, NASA Langley Research Center, Hampton, Virginia.

Schnell, T., Macuda, T., Poolman, P., Craig, G., Erdos, R., Carignan, S., Allison, R., Lenert, A., Jennings, S., Swail, C., Ellis, K., & Gubbels, A. W. (2006). Toward the "cognitive cockpit": Flight test platforms and methods for monitoring pilot mental state. In D. Schmorrow, K. Stanney, & L. Reeves (Eds.), *Foundations of augmented cognition* (2nd ed.). [author: please provide page numbers] Arlington, VA: Strategic Analysis, Inc.

Smith, P. J., McCoy, E. C., & Layton, C. (1997, May). Brittleness in the design of cooperative problem-solving systems: The effects of user performance. *IEEE Transaction on Systems, Man, and Cybernetics—Part A*, *27*(3), pp. 360 - 371.

Snow, M. P., Barker, R. A., O'Neill, K. R., Offer, B. W., & Edwards, R. E. (2006). Augmented cognition in a prototype uninhabited combat air vehicle operator console. In D. Schmorrow, K. Stanney, & L. Reeves (Eds.), *Foundations of augmented cognition* (2nd ed.) pp. 279-288. Arlington, VA: Strategic Analysis, Inc.

Steinman, A. M. (1987). Adverse effects of heat and cold on military operations: History and current solutions. *Military Medicine*, *152*, 389–392.

Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: A possible measure of fatigue. *Human Factors*, *36*, 285–297.

Takahashi, N. (2006). Research Introduction: Efficient learning algorithms for support vector machines. from http://www-kairo.csce.kyushu-u.ac.jp/~norikazu/research.en.html

Thorpe S., Fize D., & Marlot C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522.

Tremoulet, P., Barton, J., Craven, R., Gifford, A., Morizio, N., Belov, N., Stibler, K., Regli, S. H., & Thomas, M. (2006). Augmented cognition for tactical Tomahawk weapons control system operators. In D. Schmorrow, K. Stanney, & L. Reeves (Eds.), *Foundations of augmented cognition* (2nd ed.) pp. 313-318, Arlington, VA: Strategic Analysis, Inc.

U.S. Navy. (1988). *Investigation report: Formal investigation into the circumstances surrounding the downing of Iran Air Flight 655 on 3 July 1988* (Department of Defense Investigation Report). Washington, DC: Author.

Vapnik, V. (1999). *The nature of statistical learning theory*. New York: Springer-Verlag.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics, 41*, 656–669.

Verwey, W. B., & Veltman, H. A. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology-Applied, 2*(3), 270–285.

Wickens, C. D., & Hollands, J. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Wientjes, C. J. E. (1992). Respiration in psychophysiology: Methods and applications. *Biological Psychology, 34*(2–3), 179–203.

Wildervanck, C., Mulder, G., & Michon, J. A. (1978). Mapping mental load in car driving. *Ergonomics, 21*, 225–229.

Wilson, G. F., & Eggemeier, F. T. (1991). Physiological measures of workload in multi-task environments. In D. Damos (Ed.), *Multiple-task performance* (pp. 329–360). London: Taylor & Francis.

Yamada, F. (1998). Frontal midline theta rhythm and eyeblinking activity during a VDT task and a video game: Useful tools for psychophysiology in ergonomics. *Ergonomics, 41,* 678–688.