Lapses in Responsiveness:

Characteristics and Detection from the EEG

Malik T. R. Peiris, *B.E. (Hons)*

Department of Electrical and Computer Engineering

A thesis presented for the degree of

Doctor of Philosophy

University of Canterbury

Christchurch, New Zealand

January, 2008

"If we knew what it was we were doing, it would not be called research, would it?"

- Albert Einstein

Abstract

Performance lapses in occupations where public safety is paramount can have disastrous consequences, resulting in accidents with multiple fatalities. Drowsy individuals performing an active task, like driving, often cycle rapidly between periods of wake and sleep, as exhibited by cyclical variation in both EEG power spectra and task performance measures.

The aim of this project was to identify reliable physiological cues indicative of lapses, related to behavioural microsleep episodes, from the EEG, which could in turn be used to develop a real-time lapse detection (or better still, prediction) system. Additionally, the project also sought to achieve an increased understanding of the characteristics of lapses in responsiveness in normal subjects.

A study was conducted to determine EEG and/or EOG cues (if any) that expert raters use to detect lapses that occur during a psychomotor vigilance task (PVT), with the subsequent goal of using these cues to design an automated system. A previously-collected dataset comprising physiological and performance data of 10 air traffic controllers (ATCs) was used. Analysis showed that the experts were unable to detect the vast majority of lapses based on EEG and EOG cues. This suggested that, unlike automated sleep staging, an automated lapse detection system needed to identify features not generally visible in the EEG.

Limitations in the ATC dataset led to a study where more comprehensive physiological and performance data were collected from normal subjects. Fifteen non-sleep-deprived male volunteers aged 18–36 years were recruited. All performed a 1-D continuous pursuit visuomotor tracking task for 1 hour during each of two sessions that occurred between 1 and 7 weeks apart. A video camera was used to record head and facial expressions of the subject. EEG was recorded from electrodes at 16 scalp locations according to the 10-20 system at 256 Hz. Vertical and horizontal EOG was also recorded. All experimental sessions were held between 12:30 and 17:00 hours. Subjects were asked to refrain from consuming stimulants or depressants, for 4 h prior to each session.

Rate and duration were estimated for lapses identified by a tracking flat spot and/or video sleep. Fourteen of the 15 subjects had one or more lapses, with an overall rate of 39.3 ± 12.9 lapses per hour (mean ± SE) and a lapse duration of 3.4 ± 0.5 s. The study also showed that lapsing and tracking error increased during the first 30 or so min of a 1-h session, then decreased during the remaining time, despite the absence of external temporal cues. EEG spectral power was found to be higher during lapses in the delta, theta, and alpha bands, and lower in the beta, gamma, and higher bands, but correlations between changes in EEG power and lapses were low. Thus, complete lapses in responsiveness are a frequent

phenomenon in normal subjects – even when not sleep-deprived – undertaking an extended, monotonous, continuous visuomotor task. This is the first study to investigate and report on the characteristics of complete lapses of responsiveness during a continuous tracking task in non-sleep-deprived subjects. The extent to which non-sleep-deprived subjects experience complete lapses in responsiveness during normal working hours was unexpected. Such findings will be of major concern to individuals and companies in various transport sectors.

Models based on EEG power spectral features, such as power in the traditional bands and ratios between bands, were developed to detect the change of brain state during behavioural microsleeps. Several other techniques including spectral coherence and asymmetry, fractal dimension, approximate entropy, and Lempel-Ziv (LZ) complexity were also used to form detection models. Following the removal of eye blink artifacts from the EEG, the signal was transformed into z-scores relative to the baseline of the signal. An epoch length of 2 s and an overlap of 1 s (50%) between successive epochs were used for all signal processing algorithms. Principal component analysis was used to reduce redundancy in the features extracted from the 16 EEG derivations. Linear discriminant analysis was used to form individual classification models capable of detecting lapses using data from each subject. The overall detection model was formed by combining the outputs of the individual models using stacked generalization with constrained least-squares fitting used to determine the optimal meta-learner weights of the stacked system. The performance of the lapse detector was measured both in terms of its ability to detect lapse state (in 1-s epochs) and lapse events. Best performance in lapse state detection was achieved using the detector based on spectral power (SP) features (mean correlation of $\phi = 0.39 \pm 0.06$). Lapse event detection performance using SP features was moderate at best (sensitivity = 73.5%, selectivity = 25.5%). LZ complexity feature-based detector showed the highest performance ($\phi = 0.28 \pm 0.06$) out of the 3 non-linear feature-based detectors. The SP+LZ feature-based model had no improvement in performance over the detector based on SP alone, suggesting that LZ features contributed no additional information. Alpha power contributed the most to the overall SP-based detection model. Analysis showed that the lapse detection model was detecting phasic, rather than tonic, changes in the level of drowsiness.

The performance of these EEG-based lapse detection systems is modest. Further research is needed to develop more sensitive methods to extract cues from the EEG leading to devices capable of detecting and/or predicting lapses.

"Science is not about building a body of known 'facts'.
It is a method for asking awkward questions and subjecting them to a reality-check,
thus avoiding the human tendency to believe whatever makes us feel good."
- Terry Pratchett

Acknowledgements

I would like to express my gratitude to the following people and organizations for their help and support during the course of my Ph.D. study.

I cannot overstate my gratitude to my supervisors, Assoc. Prof. Richard Jones and Prof. Phil Bones. I am deeply indebted to them for having contributed greatly towards this project. Their feedback and suggestions on various aspects of the project were invaluable. I would like to thank Richard for providing me the opportunity to undertake this research project. His constant dedication, attention to detail, and encouragement have helped me immensely. I would like to especially thank Phil for his advice on technical matters and his mentorship.

I would like to thank Dr. Paul Davidson for volunteering his expertise on various technical matters related to the project, and for his suggestions on improving procedures. I would also like to thank Paul and Daniel Myall for sharing their expertise on computing and networking matters.

I am also grateful to Grant Carroll and his team in Clinical Neurophysiology for their help and support towards the continuous tracking task (CTT) study. The help given by Richard Dove and Cathy Lai for hardware and software resource support for the CTT study is also gratefully acknowledged. This study would not have taken place without the assistance of all subjects who volunteered their time and participated. I also wish to acknowledge the work of Leigh Signal, Margo van den Berg, Grant Carroll, and Phil Parkin in the air traffic controller (ATC) rating study, and would like to further thank Leigh for providing the ATC data and being willing to help on research matters.

I wish to acknowledge the support of the University of Canterbury in providing me with a Doctoral Scholarship, and the Christchurch Neurotechnology Research Programme for awarding me a Research Scholarship. I would also like to acknowledge the travel grants I received from the Christchurch Medical Research Foundation to attend the IEEE Engineering in Medicine and Biology conferences in Houston, San Francisco, and Shanghai.

I wish to thank my colleagues at Medical Physics and Bioengineering and at the Van der Veer Institute for Parkinson's and Brain Research, for their friendship, feedback, and advice.

I am eternally indebted to my parents, Mervyn and Sunetha, for instilling in me the importance of education, and for opening up windows of opportunities to me. I also wish to thank my brother Dinuk for his support and encouragement. Last but not least, I would like to thank my wife Mathina for her constant encouragement, understanding, love, and support and for proof-reading manuscripts and this thesis.

Contents

"If the brain were so simple we could understand it, we would be so simple we couldn't."

- Lyall Watson

Preface

This thesis is submitted for the degree of Doctor of Philosophy in Electrical and Electronic Engineering at the University of Canterbury. The research described herein was conducted under the supervision of Associate Professor Richard Jones and Professor Philip Bones in the Department of Medical Physics and Bioengineering, Christchurch Hospital, the Department of Electrical and Computer Engineering, University of Canterbury, and in the Van der Veer Institute for Parkinson's and Brain Research. The research presented is part of the Lapse Research Programme within the Christchurch Neurotechnology Research Programme.

The availability of a physiological and performance dataset from air traffic controllers (ATCs), collected by Dr Leigh Signal (Sleep/Wake Research Centre, University of Massey, Wellington) as part of her PhD research project "Scheduled napping on the night shift: consequences for the performance and neurophysiological alertness of air traffic controllers" (2002) was of considerable value to this project and also provided the initial impetus to the overall Lapse Research Programme.

Results of this research have been presented at a number of local and international conferences including New Zealand Physics and Engineering in Medicine (NZPEM) (Christchurch, NZ, November 2001), 26th Annual International Conference of the IEEE Engineering in Medicine and Biology (San Francisco, September 2004), NZPEM/Annual Conference of the Australasian College of Physical Scientists and Engineers in Medicine (ACPEM) (Christchurch, NZ, November 2004), 27th Annual International Conference of the IEEE Engineering in Medicine and Biology (Shanghai, China, September 2005), 28th Annual International Conference of the IEEE Engineering in Medicine and Biology (New York, USA, August 30 – September 3, 2006), and 5th World Congress of the World Federation of Sleep Research and Sleep Medicine Societies (Cairns, Australia, September 2007).

The following publications were generated during this PhD research:

Peiris, M. T. R., Jones, R. D., Carroll, G. J., and Bones, P. J. Investigation of lapses of consciousness using a tracking task: preliminary results. *Proceedings of the 26th International Conference of the IEEE Engineering in Medicine and Biology Society*, San Francisco, USA, 2004, 26: 4721-4724.

Peiris, M. T. R., Jones, R. D., Carroll, G. J., and Bones, P. J. Lapses of consciousness during a continuous tracking task (Abstract). *New Zealand Medical Journal*, 2004, Vol. 117, No 1199. http://www.nzma.org.nz/journal/117-1199/1004/.

Peiris, M. T. R., Jones, R. D., Davidson, P. R., Carroll, G. J., Signal, T. L., Parkin, P. J., van den Berg, M., and Bones, P. J. Identification of vigilance lapses using EEG/EOG by expert human raters. *Proceedings of the 27th International Conference of the IEEE Engineering in Medicine and Biology Society*, Shanghai, China, 2005, 27: 5735-5737.

Peiris, M. T. R., Jones, R. D., Davidson, P. R., Bones, P. J., and Myall, D. J. Fractal dimension of the EEG for detection of behavioural microsleeps. *Proceedings of the 27th International Conference of the IEEE Engineering in Medicine and Biology Society*, Shanghai, China, 2005, 27: 5742-5745.

Davidson, P. R., Jones, R. D., and **Peiris, M. T. R.** Detecting behavioural microsleeps using EEG and LSTM recurrent neural networks. *Proceedings of the 27th International Conference of the IEEE Engineering in Medicine and Biology Society*, Shanghai, China, 2005, 27: 5754-5757.

Peiris, M. T. R., Jones, R. D., Davidson P. R., and Bones, P. J. Detecting behavioural microsleeps from EEG spectra. *Proceedings of the 28th International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, USA, 2006, 28: 5723-5726.

Peiris, M. T. R., Jones, R. D., Davidson, P. R., Carroll, G. J., and Bones, P. J. Frequent lapses of responsiveness during an extended visuomotor tracking task in non-sleep-deprived subjects. *Journal of Sleep Research*, 2006, Vol. 15: 291-300.

Davidson, P. R., Jones, R. D. and **Peiris, M. T. R.** EEG-based Lapse Detection with High Temporal Resolution. *IEEE Transactions in Biomedical Engineering*, 2007, Vol. 54: 832-839.

Peiris M., Jones R., Bones P., Davidson P. (2007). Detection of lapses of responsiveness from spectral and non-linear features in the EEG (Abstract). *Sleep and Biological Rhythms*, 5 (Suppl 1), A40.

Glossary

ADACL	Activation-Deactivation Adjective Check List
ANN	Artificial neural network
ApEn	Approximate entropy
ATC	Air traffic controller
AUC-PR	Area under the precision-recall curve
AUC-ROC	Area under the receiver operating characteristic curve
BM	Behavioural microsleep
CMOS	Complementary Metal Oxide Semiconductor
CNS	Central nervous system
CR	Correct Rate
CTT	Continuous tracking task
DD	Deep drowsy
DIL	Driver Inattentive Level
DSS	Driver State Sensor
EC	Eyes closed
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyogram
EO	Eyes open
EOG	Electro-oculogram
ERP	Event-related potential
ESS	Epworth Sleepiness Scale
FD	Fractal dimension
FFT	Fast Fourier transform
FN	False negative
FP	False positive
ICA	Independent component analysis
IR	Infra-red
LD	Light drowsy
LDA	Linear discriminant analysis
LDS	Lane Detection System
LI	Lapse index

LS	Least-squares
LTEEG	Long-term EEG
LVQ	Learning vector quantization
LZ	Lempel-Ziv
LZC	Lempel-Ziv complexity
MSLT	Multiple Sleep Latency Test
MWT	Maintenance of Wakefulness Test
NPV	Negative predictive value
NSP	Normalized spectral power
PASS	Proximity Array Sensing System
PCA	Principal component analysis
PCs	Principal components
PERCLOS	PERcentage eye CLOSure
PPV	Positive predictive value
PR	Power ratio
PR	Precision-recall
PVT	Psychomotor vigilance task
R&K	Rechtschaffen & Kales
RBR	Relative band ratio
REM	Rapid eye movement
RMS	Root mean square
RNN	Recurrent neural network
ROC	Receiver Operating Characteristic
RT	Reaction time
SD	Standard deviation
SE	Standard error (of the mean)
SEMs	Slow eye movements
SP	Spectral power
TP	True positive
VAS	Visual Analogue Scale
WAVT	Wilkinson Auditory Vigilance Task

CHAPTER 1

Introduction

1.1 Motivation

Many manual activities involving humans, in the areas of regulation and control, have been replaced by automatic control systems. However, human operators still continue to take a critical hands-on role in much of the transportation sector. The human operator can become fatigued, lose motivation, and become considerably less effective as a controller especially during any long-term monotonous activities such as driving (Bittner *et al.*, 2000). As control function diminishes, the operator may experience periods of temporary complete losses of responsiveness (referred to as 'lapses' in this thesis). These can occur from the complex interaction of a number of factors such as boredom, physical and mental exhaustion, lack of sleep or reduced quality of sleep, and the influence of circadian rhythms (Freund *et al.*, 1995). Lapses in responsiveness can manifest themselves as brief diversions of attention, that might be described as 'mind wandering', to going to sleep ('nodding off') on the job.

A lapse prevents critical input information reaching the conscious cognitive threshold, a process that is necessary for the operator to provide an appropriate response (Brown, 1997; Freund *et al.*, 1995). Such lapses are an occupational hazard for professional transport operators such as coach and truck drivers, train drivers, air traffic controllers, and long-haul flight crew, who are expected to maintain schedules, and work shifts performing monotonous tasks for extended periods of time, despite the level of physical or mental fatigue they may be feeling. Decreases in the level of alertness can adversely affect a driver's ability to continue the assigned task safely and, as a result, the individual may find it difficult to maintain sustained attention for extended periods of time. However, the increased risk of lapses is not limited to individuals working in the transport sector. Groups such as junior

doctors who are expected to work long shifts (Barger *et al.*, 2005), nurses, police, fire fighters, and radar and sonar operators also face issues of decreased alertness and performance as a result of shift work.

Even well-rested individuals who are not sleep-deprived may experience "sleep-related states", without a preceding phase of subjectively experienced drowsiness (Sagberg, 1999). Drowsy individuals performing an active task, such as driving, often cycle rapidly between periods of wakefulness and sleep, as exhibited by cyclical variation in both EEG power spectra and task performance measures (Makeig *et al.*, 2000).

Complete loss of responsiveness (even for a few seconds) while engaged in a critical task such as driving a vehicle or landing an aircraft can have disastrous consequences in the form of serious injuries and/or multiple fatalities as well as losses to property. It is also important to recognize that the possibility of lapsing while driving is of concern to anyone using a motor vehicle and, hence, the problem is not confined to persons in the transport industry.

A device which can detect, or better still predict, impending lapses in real-time using physiological cues from an individual, would be especially beneficial to workers in the transport sector and would help minimize accidents caused by individuals lapsing while performing tasks such as driving.

1.2 Fatigue, sleepiness, and consequences of lapsing in operational environments

A number of studies have found causal relationships between sleepiness, fatigue, and lapsing. Others have quantified the impact they have on public safety in terms of the proportion of all accidents attributed to sleepiness and lapses, and the consequences of lapsing in terms of injuries, fatalities, and losses to property.

Sleepiness, like alcohol, can impair functioning at the level of the central nervous system (CNS) (Arnedt *et al.*, 2001). This can jeopardize the safety of both the sleepy driver and other road users. Studies comparing effects of fatigue and alcohol on simulated driving performance indicate that prolonged wakefulness can produce performance decrements equal to or greater than those observed at blood alcohol concentrations of 0.05% or higher (Arnedt *et al.*, 2001; Dawson and Reid, 1997; Lamond and Dawson, 1999; Maruff *et al.*, 2005; Powell *et al.*, 2001).

Sleepiness/fatigue affects critical aspects associated with driving such as reaction time, vigilance, attention, and information processing (NHTSA, 1998). Therefore, it is not entirely

surprising that researchers have found a link between an increased risk of accidents associated with sleepiness and fatigue (Horstmann *et al.*, 2000; Knipling and Wang, 1995; Masa *et al.*, 2000; NHTSA, 1998). Research has shown that even minor nightly sleep loss can have a detrimental affect on the reaction time and vigilance performance of subjects (Dinges *et al.*, 1997; Powell *et al.*, 2001).

The extent to which sleepiness and fatigue contribute to road accidents is debatable. A cross-sectional survey by Powell *et al*. (2002) found an association between self-reported sleepiness and sleep disorders, with accidents and injuries in a large sample of drivers. Horne and Reyner (1995b) found that sleep-related vehicle accidents comprised 16% of all accidents to which the police were summoned in major roads in England, and over 20% on motorways. A survey by Maycock (1997) found that 29% of drivers in the UK reported feeling close to falling asleep while driving in the previous 12 months. About 10% of all vehicle crashes in France during the 1994-8 period were related to fatigue (Philip *et al.*, 2001). A survey in Norway found that sleepiness was a contributing factor in 3.9% of all accidents (Sagberg, 1999). This factor increased markedly for night-time accidents (18.6%). They also found that one in 12 drivers admitted to having fallen asleep at the wheel in the past 12 months and about 4% of these events resulted in an accident. This equates to 1 in 250 Norwegian drivers being involved in an accident per year due to falling asleep (Sagberg, 1999). In the USA, a literature review revealed that driver sleepiness was found to be a causative factor in 1−3% of all motor vehicle crashes (Lyznicki *et al.*, 1998). However, the authors also point out that surveys of the prevalence of sleepy behaviour indicate that this estimate is conservative. This conclusion is also acknowledged by others (Knipling and Wang, 1995; Rosekind, 2005). Data from the US Fatal Accident Reporting System indicate that falling asleep accounted for between 3−4% of all fatal crashes (Pack *et al.*, 1995). The cost of sleep-related accidents in the USA is estimated to be tens of billions of dollars per year (Bittner *et al.*, 2000).

According to a review of the aviation literature by Caldwell (2005), fatigue has been estimated to be a factor in 4−7% of civil aviation mishaps and the cost of a single major civil aviation accident can exceed $500 million.

1.3 Lapse risk factors

Several risk factors have been identified as increasing the probability of a person being involved in a serious accident due to sleepiness and lapses.

Connor *et al*. (2002) demonstrated an increased risk of a crash resulting in an injury associated with (a) drivers identifying themselves as sleepy (Stanford sleepiness score[1] 4–7 *vs*. 1-3), (b) drivers reporting 5 h or less sleep in the previous 24 h, and (c) driving between 2 a.m. and 5 a.m. Other research has shown that young male adults (under 30 years old) are the most likely to have sleep-related accidents out of all age groups (Horne and Reyner, 1995a; Horne and Reyner, 1995b). People with undiagnosed or untreated sleep disorders are also at high risk (Barbe *et al.*, 1998; Findley *et al.*, 1995; Masa *et al.*, 2000), as are drivers who have taken soporific medications such as benzodiazepine anxiolytics (Barbone *et al.*, 1998; McGwin *et al.*, 2000) or sedating antihistamines (Kay and Quig, 2001). Shift workers are another group at risk of low levels of alertness due to irregular work and sleep schedules (Åkerstedt, 2003; Åkerstedt *et al.*, 2005). A survey by McCartt *et al*. (2000) found that greater daytime sleepiness, extended driving times, and a higher frequency of night-time driving were also risk factors contributing to sleep-related accidents amongst a group of truck drivers. Circadian factors have also been shown to play a major role in sleep-related accidents (Folkard, 1997; Knipling and Wang, 1995; Pack *et al.*, 1995).

1.4 Countermeasures against sleepiness and lapsing

Several strategies can be adopted to minimize the occurrence of lapses in the workplace and on the road. Getting sufficient sleep is the most effective strategy for minimizing sleepiness and fatigue and ensuring that one is sufficiently alert to carry out the task at hand (Caldwell, 2001). According to a review by Ferrara and De Gennaro (2001), the general consensus is that 7–8 hours of sleep per day is sufficient for the majority of the population. Proper shift scheduling including allowing sufficient time between work and rest periods, eliminating double shifts, shortening shift lengths, providing opportunities to sleep at the most conducive times of day, and shift rotations occurring in a clockwise direction (days to evenings to nights) are also effective countermeasures against lapsing in the operational environment (Caldwell, 2001).

[1] The Stanford Sleepiness Scale consists of seven statements spanning gradations in feelings of alertness ranging from "wide awake" to "cannot stay awake". It is easy to administer and complete, and has been validated against performance measures and in conjunction with sleep deprivation (Herscovitch, J. and Broughton, R. Sensitivity of the Stanford sleepiness scale to the effects of cumulative partial sleep deprivation and recovery oversleeping. *Sleep*, 1981, 4: 83-91).

Scheduled naps are another strategy for performance maintenance or as a recuperative function to attenuate fatigue until normal sleep is possible (Caldwell, 2001; Takahashi, 2003). Two hour naps were found to help arrest the performance decline associated with continuous work without sleep (Angus *et al.*, 1992), and even shorter naps (5–20 min) have been found to enhance performance in sleep-deprived workers (Naitoh, 1992).

Studies have shown that alertness-enhancing drugs, such as caffeine, modafinil, methylphenidate (Ritalin), and amphetamine, are quite useful for sustaining wakefulness in people deprived of sleep (Lagarde *et al.*, 1995; Penetar *et al.*, 1993; Pigeau *et al.*, 1995; Reyner and Horne, 2000).

Ensuring adequate rest breaks during work is another recommended countermeasure against fatigue (McCartt *et al.*, 2000). Use of bright light and melatonin to synchronize the body's rhythms have also been recommended (Caldwell, 2001). Although suggested by some, exercise has been found to be of little use for increasing alertness (Horne and Reyner, 1995a). Listening to music and exposure to cold air were not found to be effective countermeasures against driver fatigue (Reyner and Horne, 1998) and, if at all, their positive effect only lasted for a few minutes (Caldwell, 2001). Sharing the driving during a long trip and taking regular rest stops along the way can counter the effects of fatigue and ensure that drivers are adequately alert to drive (Cummings *et al.*, 2001). The best advice to a driver showing signs of or feeling that they are falling asleep at the wheel is to stop driving as soon as possible.

1.5 Objectives

1.5.1 Focus and goals

The focus of this thesis is on the characteristics and detection of lapses in responsiveness caused by a behavioural microsleep[2] or a temporary loss of attention during a sustained attention task. Incorporating the detection of shifts of attention into a lapse detector was deemed unachievable on the basis that there were no changes in alertness to be reflected in the EEG. This notwithstanding, the expectation from this project was that identifying reliable physiological cues indicative of lapses would contribute substantially towards the development of a lapse detection (or, better still, prediction) device which could

[2] Behavioural microsleeps (BMs) are characterized by eyelid closure, head-nodding, and markedly reduced or absent task responsiveness but may be unaccompanied by EEG indications of sleep (Ogilvie, R. D. The process of falling asleep. *Sleep Medicine Reviews*, 2001, 5: 247-270).

continuously monitor an individual in real-time. If impending lapses are detected, the system could provide a warning to the user to say that they are on the verge of having a lapse, allowing preventative or remedial action to be initiated to maintain safety.

The key goals of this project were to:

1. Conduct a review of the literature to discover and evaluate previous approaches, based upon EEG and/or EOG, used in the detection and/or prediction of drowsiness and lapses.

2. Find subject-independent features in EEG/EOG which provide reliable indications of lapses and drowsiness.

3. Investigate the efficacy of several advanced signal processing techniques as a means of substantially improving the detection accuracy of features and precursors in the brain's electrical activity and/or eye movements, thus reliably detecting drowsiness and lapses.

4. Determine the minimum number of EEG and/or EOG channels, and optimal placement of electrodes to achieve a high degree of performance (i.e., eliminate redundant information).

5. Confirm the presence and investigate the characteristics of lapses in normal non-sleep-deprived subjects.

1.6 Thesis organization

This thesis is organised into 9 chapters. The current chapter presented the motivation behind the project, provided statistics to give the reader an indication of the extent of the lapse problem in an operational context and stated the key aims of the project. Chapters 2 to 4 provide a literature review of the subject area and background information necessary to understand the key concepts discussed in this work. Chapter 2 provides a review of the biosignals and metrics used to investigate arousal and attention. Chapter 3 provides a review of the 'wake-sleep continuum'. Chapter 4 provides an overview of the state-of-art in drowsiness estimation and lapse detection and concludes by stating the research questions and hypotheses. Chapter 5 describes a rating study undertaken to determine if experts are able to detect lapses from the EEG and EOG of air traffic controllers in an operational environment. Chapter 6 describes the design, methodology, and execution of a continuous tracking task (CTT) study undertaken to collect physiological and performance data from a group of normal subjects. Chapter 7 contains analysis of data from the CTT study, focusing

on the characteristics of lapses in responsiveness. The efficacies of several linear and non-linear signal processing methods in the detection of lapses from the EEG are evaluated in Chapter 8. Finally, Chapter 9 contains the key conclusions and findings of the project, a critique of the research, and thoughts for future research in this area. The primary original contributions are presented in Chapters 5 to 8 of the thesis.

CHAPTER 2

Biosignals and Metrics for the Investigation of Arousal and Attention: A review

Chapters 2–4 provide an in-depth overview, the key concepts, and literature in areas relating to lapses. This chapter provides an introduction to biosignals associated with arousal and attention, their genesis, measurement, and their use in the identification and classification of arousal levels. Following this, a summary of the various subjective and objective measures used to assess the arousal level and sleep propensity is provided. Finally, a number of metrics used to assess task performance impairments of subjects during lowered levels of arousal are presented.

2.1 The electroencephalogram

The brain is the organ that shows the clearest changes between wakefulness and sleep (Horne, 1988). These changes manifest themselves as electrical activity in the brain, as measured on the scalp by the electroencephalogram (EEG). A discussion of cortical arousal and related brain activity would be incomplete without a brief introduction to the EEG and the various types of electrical activity that are generated in the brain (commonly referred to as "brain waves"). This section provides a summary of the key aspects of the EEG.

2.1.1 Neurophysiology

The nervous system consists of bundles of nerve cells called *neurons*. Each neuron consists of three major components − the cell body (*soma*), the receptor zone (*dendrites*), and the *axon*

which carries electrical signals from the cell body to target sites such as muscles, glands, or other neurons (Bronzino *et al.*, 2000). *Synapses* allow nerve cells to communicate with one another through axons and dendrites, converting electrical impulses into chemical signals (Carlson, 2004). A diagram of a neuron is shown in Figure 2-1.

It is generally agreed that brain function is based on the organization of the activity of large numbers of neurons into coherent groups (Bronzino *et al.*, 2000). Due to the ability of neurons in the cerebral cortex to generate and propagate electrical signals, the combined electrical activity of groups of neurons can be detected via electrodes placed on the scalp.

Electroencephalography involves the measurement, amplification, and registration of differences between fluctuating electrical field potentials as a function of time (Kamp and Lopes da Silva, 1999) and was first measured by Hans Berger in 1924 (Bronzino *et al.*, 2000). The signals recorded at the scalp primarily reflect cortical activity. Cortical activity, in turn, is affected by the electrical activity of brain structures underlying the cortex. It has been shown that cortical EEG patterns are affected by a variety of factors including biochemical, metabolic, circulatory, hormonal, neuroelectric, and behavioural factors (Bronzino *et al.*, 2000).

2.1.2 EEG recording system overview

Modern EEG recording systems are capable of recording 100 or more channels simultaneously from electrodes placed on the scalp. Figure 2-2 shows a block diagram of a single channel of an EEG system depicting the key system components.

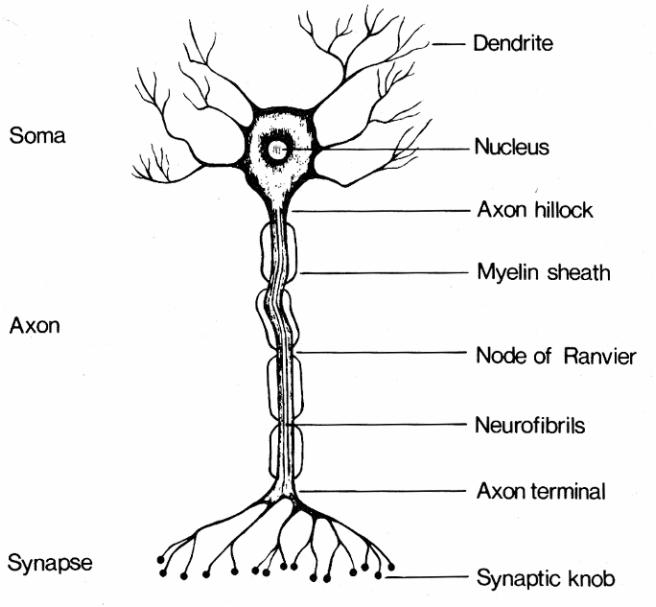

Figure 2-1 A schematic diagram of a neuron depicting its major structures (Duffy *et al.*, 1989).

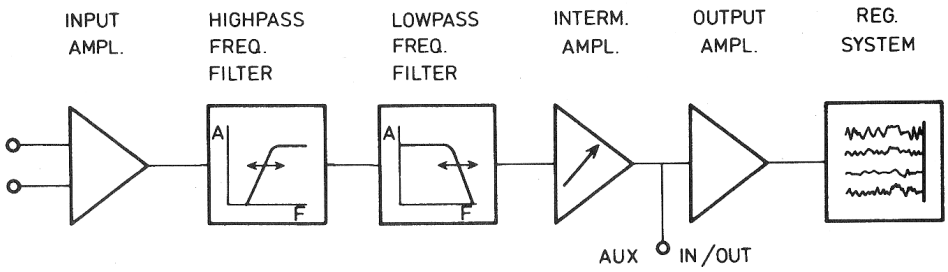

Figure 2-2 Block diagram of a single EEG recording channel (Kamp and Lopes da Silva, 1999).

The recording electrodes are connected to the input amplifier which amplifies the difference between the electrode pair while rejecting the common-mode signal components. A pair of electrodes connected to each channel (amplifier) is referred to as a *derivation*. The frequency characteristics of the signal can be shaped to the band of interest by adjusting cut-off frequencies of the high-pass and low-pass filters. The intermediate amplifier is used to maintain an adequate input voltage for the output amplifier stage, and is achieved by adjusting its gain. The output amplifier delivers the current required to drive the registration system. In the case of a digital EEG recording system, the registration system consists of an analogue-to-digital (A–D) converter. The sampled signal can be stored on disk for future analysis and displayed on screen if necessary.

2.1.3 The 10-20 International System

By placing electrodes across a wide area of the scalp, the EEG system can provide comprehensive representation of the spatial distribution of the electric potentials across the entire scalp (Kamp and Lopes da Silva, 1999).

Figure 2-3 shows an electrode map of the 10-20 International System. This system is the most widely used standard for EEG recordings and relies on standard landmarks on the skull (Duffy *et al.*, 1989). It provides adequate coverage of all parts of the head with electrode positions designed having taken underlying brain areas (frontal pole, frontal, central, parietal, occipital, and temporal) into consideration. These brain areas are shown in Figure 2-4. The electrodes are placed either 10 or 20% of the total distance between a given pair of skull landmarks and hence the term "10-20". An advantage of this system is that it can be used on a skull of any size as it does not require absolute measurements.

A specific arrangement of a group of derivations displayed simultaneously is termed a *montage*. A number of montages have been designed to make the interpretation of the EEG as easy and accurate as possible. Referential and bipolar are the two basic types of montages and they both have pros and cons associated with their use (Duffy *et al.*, 1989).

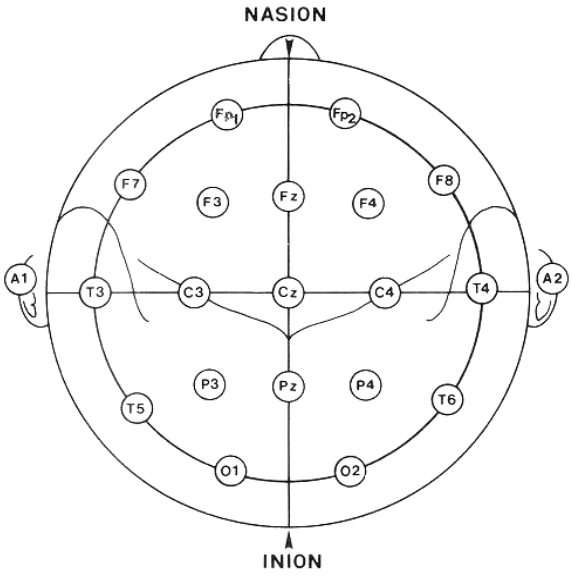

Figure 2-3 The 10-20 International System of electrode placement showing the frontal pole (Fp), frontal (F), central (C), parietal (P), temporal (T), and occipital (O) locations (Duffy *et al.*, 1989).

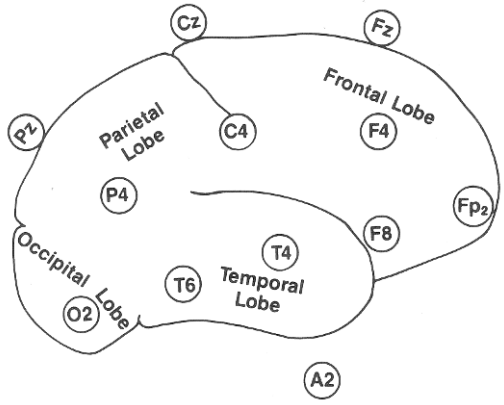

Figure 2-4 Topographic relationship between the 10–20 International System of electrode placement and areas of the cerebral cortex (Duffy *et al.*, 1989).

The referential montage uses a common reference electrode to compare the electrical activity at different electrode sites. Typically, the earlobes are used as the reference. Ideally, the common reference electrode should not be affected by any bioelectric activity, but no such ideal reference exists as the earlobes pick up temporal cerebral activity and this presents problems with localization (Duffy *et al.*, 1989). That notwithstanding, the referential montage allows valid comparisons of amplitudes in different derivations.

In contrast to the referential montage, the bipolar montage considers both electrodes to be active, and the varying voltage difference between the two electrodes is recorded. This can sometimes lead to the cancellation of the two signals at the amplifier, depending on their magnitude and phase.

2.1.4 Types of EEG activity

Cortical activity as measured by the EEG is traditionally divided into several frequency bands and is identified using the Greek letters alpha, beta, theta, and delta. These waves are specified in terms of amplitude and frequency. Their characteristics are summarized in Table 2-1, and sample tracings of these waveforms are illustrated in Figure 2-5.

Table 2-1 Characteristics of the main EEG waveforms adapted from (Duffy *et al.*, 1989; Horne, 1988; Niedermeyer, 1999).

Activity type	Amplitude/frequency/spatial characteristics
Delta	• Waves less than 3.5 Hz • High amplitude (generally over 100 μV) • Increases with deepening sleep
Theta	• Activity in the 4−8 Hz range • Amplitude around 30 μV • Does not usually occur in the normal awake adult EEG, but occurs prominently during drowsiness and sleep
Alpha	• Rhythmic activity of 8−13 Hz • Amplitude 50 μV or less. • Most prominent posteriorly • Key feature in the awake, relaxed adult. Best seen when the eyes are closed. Eye opening results in the attenuation of the alpha rhythm • The alpha amplitude and abundance may increase with attention or eye opening (instead of decreasing) following a period of drowsiness. This is termed "paradoxical alpha" (Duffy *et al.*, 1989).
Beta	• Activity ranging from 13−35 Hz • Low amplitude (<10 μV) • Best seen in anterior regions, but also commonly present posteriorly although masked by alpha activity • Occurs in cerebrum when alert or anxious

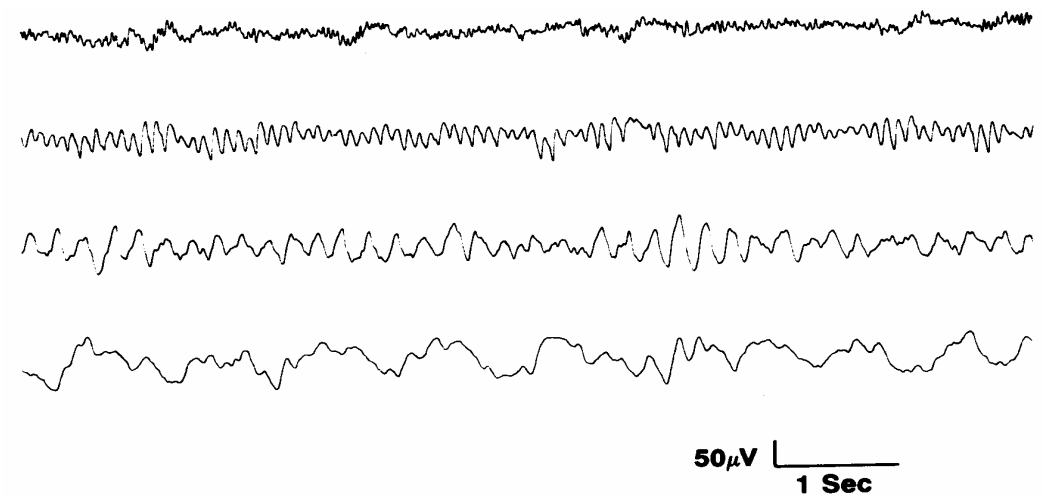

Figure 2-5 Examples of the four main types of EEG activity. From top to bottom: beta, alpha, theta, and delta activity (Duffy *et al.*, 1989).

2.2 Other relevant physiological measures

2.2.1 Eye closure

A technique known as "PERCLOS" (Wierwille and Ellsworth, 1994), which measures the percentage eye-lid closure over the pupil during a 1-min period, has been shown to be a reliable indicator of driver alertness (Mallis, 1999). This method specifically measures the proportion of time in a minute that the eyes are more than 80% closed, and uses slow eyelid closures, rather than eye blinks.

2.2.2 Eye movements

The main applications of the EOG are in ophthalmological diagnosis and in recording eye movements. EOG recordings are based on the small electro-potential difference between the front and back of the eye, with the cornea being positive with respect to the retina. Due to this reasonably constant potential difference, movements of the eyes can be measured using electrodes placed beside the eyes, with the electrode nearest to the cornea registering a positive potential and the electrode closest to the retina registering a negative potential. The EOG electrodes are often placed on the right and left outer canthi. The standard sleep scoring manual (Rechtschaffen and Kales, 1968) recommends that the right and left electrodes be placed slightly above and below the horizontal plane respectively. This allows the capture of both horizontal and vertical eye movements using a single pair of electrodes. The electrodes are usually applied to thoroughly cleaned skin using adhesive tape. Laboratories use either the same ear reference for both electrodes or contralateral references. However, if more precise eye movement measurements are required, and if both horizontal

and vertical movements are considered equally important, electrodes are placed supraorbitally and infraorbitally, in addition to placement on the outer canthi (Carskadon and Rechtschaffen, 2000). A time-constant slower than 0.3 s and a high-pass filter cut-off around 30–35 Hz are recommended when recording the EOG. Eye measurements recorded by the EOG can include saccades and smooth pursuit (vertical and horizontal), involuntary slow horizontal eye movements, and eye closure and opening.

2.2.3 Facial video

A video camera capturing a subject's face can provide valuable information regarding level of arousal. In addition to eye closure, facial video also provides information on features such as eye gaze, yawning, head droop/nodding, and facial muscle tone which can be used to estimate the level of arousal of a subject.

2.2.4 Electromyogram

Although not directly used as a tool to measure sleepiness, the electromyogram (EMG) is used to identify muscle artifacts in the EEG. The standard EMG is recorded from electrodes placed on the muscles beneath the chin. Three electrodes are placed beneath the chin, overlying the mentalis/submentalis muscles after careful skin preparation as for the EEG and EOG. The EMG is usually recorded bipolarly, with the extra electrode being used as a back-up in case of failure of the placement of the two other electrodes. The pass-band range for the EMG is in the 10–75 Hz range, with a notch filter used to remove AC interference (Carskadon and Rechtschaffen, 2000).

2.3 Subjective and objective measures of arousal

Arousal refers to a physiological state involving the level of activation of the nervous system (Segalowitz *et al.*, 1994) and is considered a "low level" process (Kahnemann, 1973). For a more detailed definition please refer to § 4.1.2. Conversely, s*leep propensity* refers to the readiness of an individual to transit from wakefulness to sleep, or the ability to remain asleep. Section 2.4 provides a brief overview of subjective and objective measures of sleep propensity.

A standard benchmark used to validate behavioural and physiological correlates of arousal states is subjective experience (Pivik, 1991). Some of the most prominent of such subjective measures are the Stanford Sleepiness Scale (Hoddes *et al.*, 1973), Visual Analogue Scale (Folstein and Luria, 1973), and the Activation-Deactivation Adjective Check List (Thayer, 1978).

The Stanford Sleepiness Scale (SSS) consists of seven statements spanning gradations in feelings of alertness ranging from "wide awake" to "cannot stay awake". It is simple to administer and complete, and has been validated against performance measures and in conjunction with sleep deprivation (Herscovitch and Broughton, 1981; Hoddes *et al.*, 1973).

The Visual Analogue Scale (VAS) measure consists of a horizontal line anchored at either end by terms characterizing the extremes of the states or moods under study. Subjects are instructed to indicate their current level of sleepiness by placing a mark on the line. The scale value is determined in arbitrary units as the distances from the left and right of the line. This scale is preferred by some investigators since subjects are less likely to remember it due to the absence of a numerical scale and hence the measure is more useful across repeated sessions (Pivik, 1991).

The Thayer Activation-Deactivation Adjective Check List (ADACL) is a self-rating list consisting of a set of adjectives describing transitory activation or arousal states and can be completed by a subject within 2 minutes. Subjects are instructed to rate how well each of the 50 adjectives describe their momentary feelings on a 4-point scale.

2.4 Subjective and objective measures of sleep propensity

The Epworth Sleepiness Scale (ESS) (Johns, 1991) is a subjective self-administered, eight-item questionnaire that has been validated and found to be a simple and reliable method for measuring persistent daytime sleepiness in adults (Johns, 1992). It provides a probability measure of falling asleep in a variety of situations.

Two objective measures of sleep propensity are the Multiple Sleep Latency Test (MSLT) (Thorpy, 1992) and the Maintenance of Wakefulness Test (MWT) (Mitler *et al.*, 1982).

The MSLT is one of the most validated and accepted objective tests of sleepiness in clinical practice (Cluydts *et al.*, 2002). Subjects are instructed to lie in a quiet darkened room and are encouraged to fall asleep while their EEG and EOG are recorded. The test consists of a series of four to five nap opportunities spaced at 2-hourly intervals, with termination of each nap after either 20 min of wakefulness or 15 min of sleep. Sleep onset time is defined as the time required to reach the first epoch scored as sleep according to the standard criteria (Rechtschaffen and Kales, 1968). The sleep latency value is recorded as the mean of the sleep latencies of all naps, and is generally interpreted as: mean latency <5 min signifies pathological sleepiness, >10 min is considered normal, and between 5 and 10 min is indicative of indeterminate sleepiness (Carskadon *et al.*, 1986).

The methodology of the MWT is similar to that of the MSLT, except that the subjects are instructed to attempt to stay awake, sitting in a dark room for a period between 20–40 min, without taking extraordinary steps to remain awake. This test was designed, in part, to provide a more sensitive and clinically valid measure of sleep propensity than the MSLT (Shen *et al.*, 2006). The MWT has proven to be better than the MSLT in assessing cases which require sustained attention, since its scores have correlated better with the ability of subjects to resist sleep in monotonous circumstances (Cluydts *et al.*, 2002).

Although both the MSLT and MWT relate to 'sleep propensity', they measure different attributes. The MSLT is primarily felt to measure physiological sleepiness, akin to the primary sleep drive, whereas MWT measures alertness (or 'wakefulness') presumably akin to the primary wake (arousal) drive (Shen *et al.*, 2006).

2.5 Measures of task performance

There is considerable evidence to suggest that subjects with low levels of cortical arousal have impaired perceptual skills, reasoning ability, judgement and decision-making capabilities (Bonnet, 1994; Dinges and Kribbs, 1991). This effect is evident from an increase of response omissions (lapses), general cognitive slowing, memory deficits, and increased reaction times (Curcio *et al.*, 2001). Performance tasks which can be used to assess these fluctuations can be divided into two broad categories − psychomotor and cognitive tests.

Types of psychomotor tasks include auditory and visual reaction time (simple or multiple choice), tracking, and tapping tests. The range of cognitive tests includes attentional, memory, and logical reasoning tasks. As psychomotor tasks are used and referred to in this thesis, a brief description of the key auditory and visual psychomotor task tests that have been used in vigilance research is outlined next.

2.5.1 Psychomotor tasks – auditory

One of the common tests used to assess auditory vigilance is the *Wilkinson Auditory Vigilance Task (WAVT)* (Wilkinson, 1970). Standard tones (500 ms duration) are presented every 2 s in a background of high intensity (85 dB) white-noise during the 60-min test. Forty target tones of 400 ms duration are also presented randomly during the test. The subject is required to distinguish between the target and standard tones and respond to the target tone using a push button. Performance is evaluated in terms of hits, omissions, false positives, and reaction time to target tones. The WAVT has been shown to be sensitive to vigilance fluctuations caused by sleep deprivation. It has also been reported to be "long and boring" and require "cumbersome experimental apparatus" (Curcio *et al.*, 2001).

Some of the other auditory vigilance tasks mentioned in the literature include the *Simple Reaction Time Test* (Glenville *et al.*, 1978), where subjects have to respond as soon as possible to all presented stimuli presented at random intervals over a 10-min period, a passive sonar detection task (Makeig and Inlow, 1993), and a 23-min continuous auditory performance task (Arruda *et al.*, 1999).

2.5.2 Psychomotor tasks – visual

The *psychomotor vigilance task (PVT)* is one of the most widely used tests to measure visual psychomotor performance. The PVT was developed by Wilkinson and Houghton (1982) and analyses were developed by Dinges *et al*. (1985). It is sold commercially by Ambulatory Monitoring Inc. (Ardsley, NY, USA). It was developed to track the time-course of the dynamic changes induced by the interaction of the homeostatic drive for sleep and the endogenous circadian pacemaker (Dorrian *et al.*, 2005).

It has been shown that RT performance over a relatively short time period can reveal changes in performance caused by fatigue and drugs and, therefore, serve as a useful indicator of general psychomotor impairment (Powell, 1999). The PVT involves a simple reaction-time task (as opposed to a choice RT) and involves a combination of prefrontal cortex executive attention and traditional stimulus-response testing (Dorrian *et al.*, 2005). The designers of the PVT deliberately avoided choice RT to "minimize continued learning and strategy shifts that can occur even in four-choice RT tasks" (Dorrian *et al.*, 2005) and have shown that the PVT has minimal learning effects (Dinges *et al.*, 1997).

The PVT requires individuals to respond as fast as possible to the presentation of digits on an LED digital counter (bright red light stimulus) by pressing a response button. This stops the stimulus counter and displays the RT in milliseconds for a 1-s period. The subjects are instructed to press the button as soon as the stimulus appears. The test is typically 10 min in duration and the inter-stimulus interval is 2–10 s. This results in approximately 90 RTs per test which is considered a relatively high signal load. Reaction times exceeding 500 ms in the PVT are arbitrarily defined as "lapses" (Dorrian *et al.*, 2005).

The PVT is generally considered to be a validated, reliable and sensitive test of vigilance and simple visual reaction time (Dinges *et al.*, 1998) and has been shown to be sensitive to sleep deprivation (Dorrian *et al.*, 2005). Due to the discrete nature of the PVT stimuli, there is no continuous sampling of the subject's performance, and the nature of their arousal and attention levels is unknown during these inter-stimulus periods.

Comptrack is a visual psychomotor task designed by Makeig and Jolley (1996). This 2-D compensatory tracking task requires subjects to manipulate a trackball to produce forces

(proportional to velocity) countering quasi-random forces that tend to "blow" a circular disk off an invisible "slippery hill" at the center of the screen. The apex of the screen is marked by a target ring. Subjects use trackball movements to maintain the disk as near as possible to the ring. Alert task performance (mean distance <3 disk radii) requires the user to make appropriate trackball movements on average at least once per second. When no user input is generated, the unseen mound which generates the surface force keeps the disk from remaining within 3 disk radii of the target bull's-eye for 95% of the time. Differentiation between "alert" and "drowsy" (or absent performance) is achieved by distinguishing between times when the target "escapes" beyond 3 disk radii versus times when the subject is able to keep the target near to his or her best training performance level.

Other visual tasks used are the *Four-Choice Reaction Test* (Wilkinson and Houghton, 1975) – a complex visual reaction time task where the subject has to respond to one of four lights by pressing the corresponding button – and the *Simulated Assembly Line Task* (Walsh *et al.*, 1992), where the subject is instructed to "repair" damages to a circuit board presented on a screen.

2.5.3 Dual-task paradigms

The dual-task paradigm assesses the ability of subjects to perform two tasks simultaneously and respond appropriately to both tasks. As the dual-task places additional burden on the subject over that of a single task, subjects may find it more stimulating and this may cause an increase in the arousal level (Verwey and Zaidel, 1999; Wierwille and Ellsworth, 1994). This notwithstanding, a dual-task paradigm (i.e., driving simulator plus subtraction task) was found be sensitive to impairments in task performance caused by moderate sleep loss, which, somewhat surprisingly, was not demonstrated on the PVT (Rupp *et al.*, 2004). Dual tasks have also been used in combination with other performance measures in assessing changes of vigilance levels caused by caffeine and alcohol (Brice and Smith, 2002; Williamson *et al.*, 2001).

2.5.4 Driving simulators

Driving simulators consist of varying degrees of complexity and realism, and have been used to investigate the consequences of a variety of conditions, including sleep deprivation (Fairclough and Graham, 1999; Lenne *et al.*, 1998) and drowsiness estimation (Lin *et al.*, 2006; Lin *et al.*, 2005a). They allow the measurement of a number of driving-related skills, such as vehicle control (e.g., steering and pedal control), lane position (tracking), and speed deviation. Simulators can provide valuable objective information regarding a person's ability to safely operate a motor vehicle during periods of diminished arousal in a more realistic experimental setting.

CHAPTER 3

Wake-Sleep Continuum: A Review

The wake-sleep continuum has fascinated researchers for many years. There has been much interest in finding cortical, behavioural, and performance changes that occur in human subjects as they traverse along this continuum. Lindsley (1952) presented a succinct summary of the relationship between EEG measures, levels of awareness, and behavioural efficiency which has been widely accepted and also shown to be experimentally useful (reproduced in Table 3-1). This chapter provides a review of wakefulness, sleep, and the transitional phase between these two fundamental states.

3.1 Wakefulness

The most prominent feature of the awake-EEG is the posteriorly-dominant alpha rhythm (Duffy *et al.*, 1989). This is most visible during relaxed wakefulness with eyes closed and attenuates with attention and when the eyes are opened (Carskadon and Rechtschaffen, 2000). When the eyes are opened, the EEG activity becomes desynchronized (also referred to as "mixed frequency" as opposed to narrow-band activity) and of low amplitude (Empson, 1986). The EEG of a highly attentive person whose eyes are open contains a predominance of EEG activity in the beta range (15–25 Hz) (Makeig *et al.*, 2000). The EMG during wakefulness displays a high level of tonic activity.

Frequent, gaze-related, and fast eye movements in any direction (saccades) are a common eye movement during wakefulness (Santamaria and Chiappa, 1987). Two other types of eye movements termed *blinks* and *mini-blinks* are also observable during this state.

Table 3-1 Psychological states and their EEG, level of awareness, and behavioural correlates. Source: Lindsley (1952) as adapted by Empson (1986). Loosely defined, 'synchronization' refers to the adjustment of EEG activity in various brain regions to work in unison in the time and frequency domains. 'Desynchronization' refers to the loss of synchronization between various cortical regions.

Level of consciousness/arousal	EEG	State of awareness	Behavioural efficiency
Strong, excited emotion (fear, rage, anxiety)	Desynchronized; low to moderate amplitude; fast mixed	Restricted awareness; divided attention	Poor (lack of control, freezing-up, disorganised)
Alert attentiveness	Partially synchronized; mainly fast, low-amplitude waves	Selective attention, but may vary or shift. 'Concentration'.	Good (efficient, selective, quick reactions)
Relaxed wakefulness	Synchronized; optimal alpha rhythm	Attention wanders – not forced. Favours free association	Good (routine reactions and creative thought)
Drowsiness	Reduced alpha and occasional low-amplitude slow waves	Borderline, partial awareness. Imagery and reverie	Poor (uncoordinated, sporadic, lacking timing)
Light sleep	Spindle bursts and slow waves. Loss of alpha	Markedly reduced consciousness (loss of consciousness). Dream state	Absent
Deep sleep	Large and very slow waves (synchrony but on slow time base)	Complete loss of awareness (no memory for stimulation or for dreams)	Absent
Coma	Iso-electric to irregular large waves	Complete loss of consciousness, little or no response to stimulation. Amnesia	Absent
Death	Iso-electric: gradual and permanent loss of EEG activity	Complete loss of awareness as death ensues	Absent

When the eyes are open, large amplitude, vertical eye movements called blinks are observable in the EOG. The duration of a non-conscious blink is approximately 250 ms, and in most cases are incomplete (eye-lids do not touch) (Doane, 1980). The typical eye-lid excursion is around 8–9 mm with a peak velocity ranging 16–19 cm/s (Doane, 1980). Bell's

phenomenon[3] does not typically occur during normal non-conscious fast blinks (Doane, 1980). However, retraction of the eye globe in the anterior-posterior direction (1–6 mm) occurs due to the considerable pressure exerted on the anterior surface of the globe by the descending upper lid (Doane, 1980).

During eye closure, blinks of small to moderate amplitude ('mini blinks') are recorded simultaneously with posterior EEG alpha activity. These appear as rapid deflections in the EOG, reaching a peak height after 50 to 100 ms, and with maximum event duration of around 400 ms. It has been shown that each subject has a dominant mini-blink frequency (Santamaria and Chiappa, 1987).

3.2 Sleep

On average, humans spend approximately a third of their life sleeping (Sejnowski and Destexhe, 2000) and it is accepted that sleep is a complex mixture of physiological and behavioural processes. Sleep is generally identified through a set of characteristic behaviours: (a) a typical body posture (for a given individual), (b) physical inactivity, (c) elevated arousal threshold, (d) state reversibility with stimulation (i.e., the subject can transition from sleep to wake state given sufficient stimulation), (e) regular occurrence influenced by a circadian clock, and (f) closed eyes (Carskadon and Dement, 2000; Horne, 1988; Ogilvie, 2001). Carskadon and Dement (2000) defined sleep succinctly as a "reversible behavioural state of perceptual disengagement from and unresponsiveness to the environment". It has been suggested that the process of falling asleep can be optimally measured by considering a convergence of behavioural, EEG, physiological, and subjective information (Ogilvie, 2001).

3.2.1 Sleep stages

Sleep follows a regular cycle each night in healthy individuals, and the EEG pattern changes in a predictable way several times during sleep. The standard criteria used to define sleep onset and subsequent sleep stages depend on the presence of specified patterns of physiological activity. These criteria were developed by Rechtschaffen and Kales (1968) and

[3] Bell's phenomenon is a normal defence reflex present in about 75% of the population, resulting in upward movement of the eyes when blinking forcefully or when threatened by a foreign object. (Jones, D. H. Bell's phenomenon should not be regarded as pathognomonic sign. *British Medical Journal*, 2001, 323: 935).

their rating scale (often referred to as 'R&K criteria') is widely accepted to be the gold standard for sleep scoring.

When scoring sleep stages, the entire record is typically divided into epochs of 20 or 30 s duration (Carskadon and Rechtschaffen, 2000) and each epoch is scored using standard criteria (Rechtschaffen and Kales, 1968) and assigned a single sleep stage. These criteria are listed in Table 3-2 in terms of EEG, EOG, and EMG features. Examples of the EEG signatures used to identify each sleep stage are illustrated in Figure 3-1. Where more than one stage is present in an epoch, the stage which takes up the greater portion of the epoch is scored as the stage of that epoch. According to the R&K standard, sleep is broadly divided into two main states − non-rapid eye movement (non-REM) sleep and rapid eye movement (REM) sleep.

Non-REM sleep is traditionally divided into 4 sub-stages (stages 1, 2, 3, and 4) and indicates an increase in sleep depth with sleep stage, with arousal threshold increasing as one traverses from stage 1 to 4. The brain is relatively inactive during the non-REM stages and muscle tone is maintained throughout and evident by low amplitude EMG activity. The EEG is characterized by synchronous EEG with features such as sleep spindles, K-complexes, and high voltage slow waves. *Sleep spindles* are bursts of rhythmic activity between 11–15 Hz with variable amplitude and duration and occur in a widespread distribution. The *K-complex* is a slow-wave transient and is commonly diphasic and its amplitude is generally maximal at the vertex. K-complexes are large amplitude waves with amplitudes as high as several hundred microvolts. A sleep spindle may immediately follow a K-complex. A K-complex may typically last for 1 s or longer (Duffy *et al.*, 1989). An example of a sleep spindle and K-complex is shown in Figure 3-1 in the stage 2 sleep EEG trace.

REM sleep, in contrast, is signified by an electrically active brain (as seen on the EEG), a paralysed body, and episodic bursts of rapid eye movements. Unlike non-REM sleep, REM sleep is not divided into sub-stages. The background activity during REM sleep is paradoxically similar to that observed during wakefulness with the eyes open (Duffy *et al.*, 1989).

3.2.2 The sleep cycle

The following description is predominantly based on a summary of normal human sleep presented by Carskadon (2000). The normal human adult enters sleep through the non-REM state. REM sleep occurs approximately 80 minutes or longer after sleep is entered. Sleep patterns then cycle between non-REM and REM, with a period of about 90 min.

Table 3-2 Modified Rechtschaffen and Kales sleep-staging scale adapted from (Carskadon and Rechtschaffen, 2000; Empson, 1986).

Stage/state	EEG	EOG	EMG
Relaxed wakefulness	*Eyes closed:* rhythmic alpha; prominent in occipital derivation; attenuates with attention *Eyes open:* relatively low voltage, mixed frequency	Voluntary control; saccadic movements; blinks	Relatively high tonic activity; voluntary movement
Stage 1	Relatively low voltage, mixed frequency; theta activity may occur with greater amplitude, vertex sharp waves	Involuntary, slow, rolling eye movements	Moderate tonic activity - may be slightly reduced from waking state
Stage 2	*Background:* relatively low voltage, mixed frequency *Sleep spindles:* waxing, waning, 12–14 Hz (≥0.5 s) *K-complex:* negative sharp wave followed immediately by slower positive component (≥0.5 s); spindles may ride on K complexes; K complexes maximal in vertex; spontaneous or in response to sound	Occasional slow eye movements near sleep onset	Moderate tonic activity
Stage 3	≤50% high amplitude (≥75 μV), slow frequency (≤2 Hz); frontally maximal	None, picks up frontal EEG	Low tonic activity
Stage 4	>50% high amplitude, slow frequency	None, picks up frontal EEG	Low or very low tonic level
REM	Relatively low voltage, mixed frequency; saw-tooth waves; theta activity; slow alpha	Phasic rapid eye movements	Tonic suppression; phasic twitches

The first cycle of sleep in the normal young adult begins with stage 1 which can last approximately 1 to 7 min. Stage 1 sleep does not constitute true sleep as a person is not fully asleep but merely drowsy during this period. Stage 1 sleep occurs as a transitional stage throughout the night.

Stage 2 of non-REM sleep follows stage 1, and is identified in the EEG by the presence of sleep spindles and K-complexes (see Figure 3-1). It is generally accepted that true sleep commences at the onset of stage 2 (Ogilvie, 2001). Stage 2 lasts for between 10 and 25 min. Stage 2 falls into the 'light sleep' category and a person can be relatively easily awakened from stage 2 sleep.

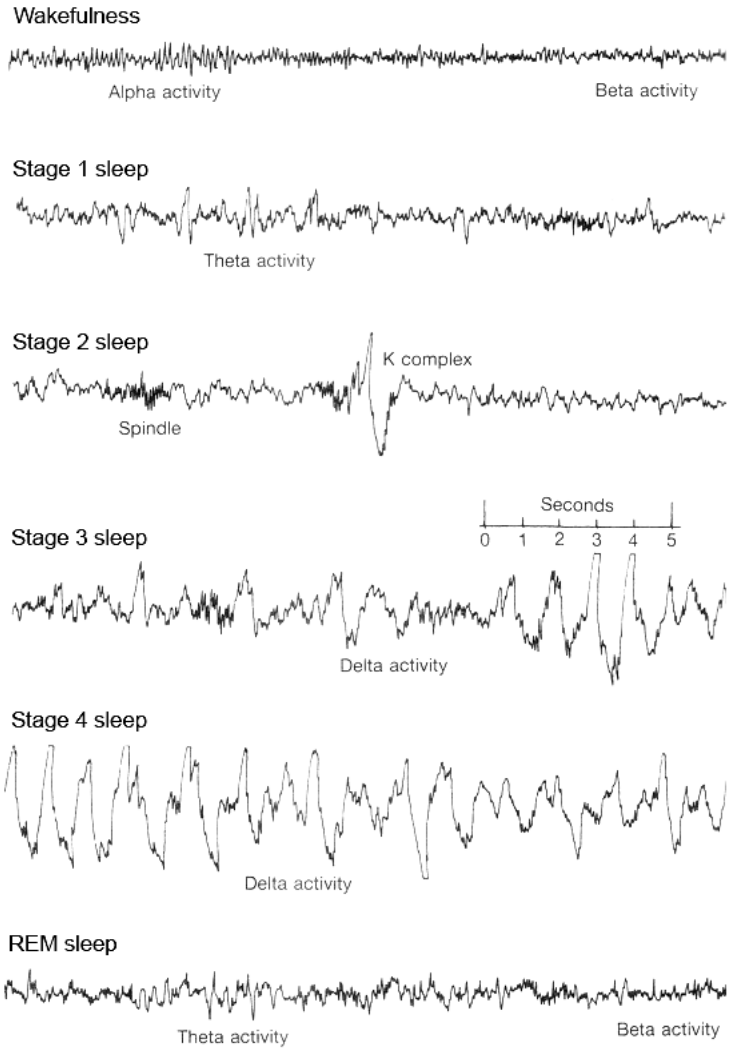

Figure 3-1 EEG of human sleep stages. From top to bottom, the tracings show typical EEG features for each of the stages from wakefulness to REM sleep (Horne, 1988).

As stage 2 progresses, there is a gradual appearance of high-amplitude slow-wave activity in the EEG. When an epoch contains between 20 and 50% of such slow wave activity, it is classified as stage 3 sleep. Stage 3 is only a transitional stage into stage 4 and only lasts a few minutes in the first sleep cycle.

When the record contains more than 50% of high-amplitude slow wave activity, it is classified as stage 4 sleep. This lasts 20–40 min in the first cycle. A much larger stimulus is required to achieve arousal during stages 3 and 4 (cf. stages 1 and 2). Stages 3 and 4 are also referred to as 'slow wave sleep' or 'deep sleep' in the literature.

A series of body movements indicate a transition into lighter non-REM sleep stages. There may be a brief episode (1–2 min) of stage 3 sleep, followed by 5–10 min of stage 2 sleep interrupted by body movements prior to entering the initial REM episode. The first REM

cycle lasts 1–5 min. The arousal threshold during REM sleep is variable throughout the night.

Non-REM and REM sleep episodes (excluding stage 1) continue to alternate through the night in a cyclical manner. REM sleep cycles generally get longer across the night, and stage 3 and 4 sleep episodes tend to get shorter and may even disappear altogether across the night as stage 2 increases and dominates the non-REM portion of sleep. The average non-REM/REM cycle duration across the night is 90–110 min. Approximately 75–80% of nocturnal sleep is non-REM.

3.2.3 A snapshot of the neurobiology of sleep

At sleep onset, brief episodes of 7–14 Hz synchronized spindles occur in the thalamus and cortex, producing large-scale spatio-temporal coherence throughout the forebrain. During sleep, the low-amplitude high-frequency activity in the neocortex characteristic of the wakeful state is replaced with high-amplitude low-frequency rhythms (Steriade *et al.*, 1993). The cortex alternates between periods of slow-wave sleep in the 2–4 Hz range and periods of rapid eye movement (REM) sleep, characterized by sharp waves of activity in the pons, the thalamus, and the occipital cortex.

3.2.4 Sleep deprivation and its physiological and behavioural effects

Research has shown that sleep is essential to maintain normal cerebrocortical function in humans. Over 50 sleep-deprivation studies have shown that the human brain (specifically the cerebral cortex) exhibits the greatest sensitivity to sleep loss (Horne, 1988). It has also been shown that the rest of the body is relatively unaffected by sleep deprivation (Kleitman, 1963) and that sleep is not necessary for the restitution of the rest of the body. Rather, it is physical rest that facilitates body restorative processes, and this is achieved during relaxed wakefulness in humans (Horne, 1988).

Horne and Pettitt (1985) found that loss of motivation was the main cause of performance impairment during initial sleep loss. They found that people could maintain baseline performance levels during an auditory vigilance task after 36 hours of sleep deprivation. However, if the subjects were further sleep deprived, they also found that there was some form of cerebral impairment causing performance impairment. They also reported that only a portion of the lost sleep needs to be recovered for the subject to return to normal cerebral function.

Horne divided sleep into two parts, called *core sleep* and *optional sleep* (Horne, 1988). He stated that core sleepiness, which he speculated was caused by cerebral "wear and tear",

needed core sleep for its restitution. During a normal night's sleep, core sleep occupies the initial 4–5 hours. The amount of stage 4 sleep following sleep deprivation was found to be influenced by the length of prior wakefulness, and to this extent best fits any restorative role. Horne also suggested that performance and behavioural impairments as indicated by various psychological tests in sleep deprived subjects are partly due to the build-up of some sort of behavioural drive to sleep. He referred to this as "optional sleepiness" and stated that it can be counteracted to a large extent by the sleep-deprived subject applying more effort. He speculated that the role of optional sleep may be to conserve energy, particularly in smaller mammals who are unable to exhibit relaxed wakefulness, and may also be a time filler when the mammal has little else to do, for example during darkness.

3.3 Wake-sleep transition

As mentioned at the start of this chapter, the wake-sleep transition is the period between wakefulness and unequivocal sleep, where unequivocal or true sleep is traditionally considered to have been reached when standard stage 2 sleep is present (Hori *et al.*, 1994). Early research tended to focus on sleep and ignored the wake-to-sleep transition to a large extent. However, from a key perspective of this project – namely, detecting behavioural microsleeps and lapses in responsiveness – familiarity with the wake-sleep transition and the EEG and EOG changes that accompany it is desirable since off-peak performance, slowing of cognitive functions, including increased reaction times, has been associated with, and suggested as, early behavioural signs of this transitional period (Ogilvie, 2001).

As presented later in this section, there is support for the contention that this transitional period is a unique period that cannot be definitively classified as either wakefulness or sleep (Hori *et al.*, 1994). It is important to emphasize that research has found no exact sleep-onset point, but rather a sleep-onset *period* during which there is a gradual transition from wakefulness to sleep (Rechtschaffen, 1994).

A review of the literature by Ogilvie (2001), incorporating metrics such as active behavioural responses, several EEG measures, ERP components and other physiological indicators, and evidence of mental activity, also suggested that there is no "moment" or "point" when sleep begins. Rather, when superimposed, the metrics indicated gradual changes during the wake-sleep transition.

The conceptual confusion surrounding this transitional period is reflected in the variety of terminology and definitions used to describe it. Kleitman (1963) used terms such as "sleepiness", "drowsiness", "languor", "inertness", "heaviness of eyelids", and "sluggishness" interchangeably to refer to this phenomenon and suggested phrases such as

"onset of sleep", "falling asleep", "going to sleep", "dormition and hypnagogic-state" to label this transitional period. Santamaria and Chiappa (1987) preferred to use the term "drowsiness" to refer to this period, while Ogilvie (1984) labelled it the "sleep onset period". According to R&K criteria, this transition equates to "stage 1 sleep" (Rechtschaffen and Kales, 1968).

This transitional period has been marked by features such as dreamlike thinking, poor agreement between standard EEG stages of sleep and the subjective experience of being asleep, and discrepancies between the EEG stages of sleep and behavioural responses (Hori *et al.*, 1994).

3.3.1 EEG changes

In general, the transition from wakefulness to sleep can be described in terms of the EEG as a shift from desynchronized to more synchronized EEG patterns (Ogilvie, 2001). This is believed to occur "when excitatory and inhibitory post-synaptic potentials become activated with increasing simultaneity, resulting in the appearance of an EEG containing a more limited variety of frequencies" and, hence, increases in amplitude (Ogilvie, 2001).

A number of studies have reported decreases in alpha activity accompanied by increases in theta activity during the transition from wakefulness to sleep (Davis *et al.*, 1937; Davis *et al.*, 1938; Dement and Kleitman, 1957; Makeig and Inlow, 1993; Rechtschaffen and Kales, 1968). In addition, Makeig and Jung (1995) found a decrease in beta activity accompanying the decrease of alpha and increase of theta activity during the transition.

Kooi *et al*. (1964) investigated the EEG patterns of the temporal region in a study involving 218 normal adults, and found that drowsiness facilitated the appearance of temporal EEG asymmetries (especially theta and delta transients), increasing from 50.8% for fully alert individuals to 71.4% during drowsiness. They also found that these asymmetries had a fixed time relationship to alpha disappearance and "should be considered one of the normal events accompanying the reorganization of brain rhythms concomitant with lowering of vigilance from relaxed wakefulness to drowsiness". They also found that the temporal transients subsided in conjunction with alpha disappearance and found a correlation between the presence of these patterns, a "poorly modulated alpha rhythm", and prominence of central beta. The left temporal area was involved 3–4 times more often than the right.

Maulsby *et al*. (1968) conducted a large study involving 200 male US Air Force flight personnel, all of whom had awake alpha activity. 18 EEG channels (fixed bipolar montage), horizontal and vertical EOG, and other parameters were recorded. Drowsiness and sleep

were found to have "infinitely more complex and variable patterns than the wakeful EEG pattern". The authors believed that the appearances of one or more of the following signs were found to be the earliest indicators of drowsiness:

1. Decrease in alpha amplitude, usually over a period of 10–30 s (seen as the first sign in 10% of subjects)
2. Appearance of theta or increase in its amplitude (seen in 8% of subjects)

Santamaria and Chiappa (1987) provided a detailed literature review and description of the changes that occur in the EEG, and eye movements during the wake-sleep transition in their landmark paper. They noted a shift in the awake, posterior predominant EEG activity (40–80 μV, 9–10 Hz) to generalized, low-amplitude, 3–6 Hz activity. They considered the neurophysiological basis for this shift in EEG to be unknown and the electrophysiology during the wake-sleep transition to be complex, with an individual exhibiting different patterns each time they traverse this continuum.

A study by Torsvall and Åkerstedt (1987), which recorded EEG and EOG of train drivers during night time journeys reported that simultaneous appearance of slow eye movements (SEMs[4]) and alpha bursts were characteristic of episodes of the driver "dozing off". They found that self-rated sleepiness increased sharply as the night progressed, along with spectral power in the alpha band and SEMs. They also observed an increase (to a lesser degree) in the theta and delta bands. Torsvall and Åkerstedt (1988) also found that SEM activity, delta, and theta power densities peaked immediately prior to 'dozing off' events, whereas alpha power density peaked approximately during the last 21 s preceding 'dozing off' events.

Cajochen *et al*. (1995) studied the power density changes in the EEG during a 40-hour period of sustained wakefulness in 9 healthy females under constant controlled conditions. They found a monotonic increase in theta and alpha frequency power (6.25–9.0 Hz) during 40 hours of wakefulness. Since the subjects were prevented from falling asleep, there was no observation of increase in power in the delta band. This contrasts with Torsvall and Åkerstedt's (1987) study which reported an increase in delta power with prolonged wakefulness in train drivers who were carrying out their job at night.

Badia *et al*. (1994) found that analysis of the EEG in 1-Hz bins using epochs of 5-s duration provided additional information regarding the changes that occur in the brain during the wake-sleep transition. The majority of subjects showed the largest increases in brain activity

[4] See § 3.3.2 for description.

at 3 and 4 Hz and the largest decreases at 10 Hz. Importantly, they found that the transition from wakefulness to sleep is not smooth but, rather, consists of frequent oscillations between wakefulness and sleep in all subjects. Their results are in agreement with changes in the traditional EEG bands (alpha, beta, and theta) reported elsewhere (Davis *et al.*, 1937; Davis *et al.*, 1938; Dement and Kleitman, 1957; Makeig and Inlow, 1993; Rechtschaffen and Kales, 1968). Badia *et al*. (1994) also observed that sleep or wake activity did not occur concurrently at all EEG measuring sites. There were inter-subject differences in the specific EEG frequency showing the largest EEG power change and the site on which it was measured.

De Gennaro *et al*. (2001) showed frequency-specific topographical changes and different timings of sleep onset across different brain sites during the wake-sleep transition. They found that these changes are distributed along the antero-posterior axis, with no inter-hemispheric differences. This finding confirmed the belief that sleep does not necessarily begin simultaneously in all cortical areas and that posterior regions of the brain are the last to show EEG changes from wakefulness to sleep (Wright *et al.*, 1995).

3.3.2 Eye movements

Several types of eye movements have been found to correlate with the wake-sleep transition. Santamaria and Chiappa (1987) observed (a) an increase in the frequency of eye movements and a decrease in their velocity and excursion, (b) the disappearance of blinks and mini-blinks to be the earliest reliable sign of the transitional period from wakefulness to sleep, and (c) two-thirds of subjects spent varying amounts of time with no eye movements, with the remaining one-third showing small, high-velocity (<100 ms), and irregularly repetitive (5–30 /s) eye movements.

An important type of eye movement associated with the wake-sleep transition is SEMs. These are moderate-to-large amplitude, slow (average duration 3–4 s), sinusoidal, and horizontal eye movements which increase in amplitude with increasing drowsiness (Kojima *et al.*, 1981; Kuhlo and Lehmann, 1964; Liberson and Liberson, 1965; Maulsby *et al.*, 1968; Santamaria and Chiappa, 1987; Torsvall and Åkerstedt, 1987; Torsvall and Åkerstedt, 1988). Reasonable correlations have been observed between SEMs and EEG throughout the wake-sleep transition (De Gennaro *et al.*, 2000; Santamaria and Chiappa, 1987). De Gennaro *et al*. (2000) quantified this correlation via multiple regression, reporting a multiple R equal to 0.84 using all EEG bands as predictors and SEM percentage as a criterion. SEMs have been reported to occur during an awake eyes-closed EEG but observed to become larger and more regularly sinusoidal with simultaneous EEG slowing or decrease in amplitude, and disappear with arousal and deeper sleep (Kuhlo and Lehmann, 1964). Although SEMs are typically observed with eyes closed, the presence of SEMs with eyes open has also been

reported (Åkerstedt and Gillberg, 1990). In a study using SEM as a secondary indicator of sleep onset, Ogilvie *et al*. (1988) found SEMs to be absent near the start of the wake-sleep transition when subjects were able to respond rapidly to tone stimuli. They observed that the occurrence of SEMs peaked in association with mean reaction times and then disappeared when subject responses ceased at the onset of true sleep. This corresponded to stage 5 in the Hori scale (see § 3.3.4). The peaking of SEM activity up to sleep onset and the subsequent decrease after sleep onset has also been observed by De Gennaro *et al*. (2000). They also observed that the spectral power in the sigma band (12–14 Hz) of the EEG was the best predictor of SEM variations. They suggest that since the sigma band overlaps the frequency of sleep spindles, that sleep spindles could trigger the reduction and final disappearance of SEMs as stage 2 sleep is reached.

Liberson and Liberson (1965) observed SEMs in most subjects during drowsiness during afternoon naps. These started before alpha disappearance (5–15 s before in 80%, up to 1 min before in some), increased during the first 15 s of drowsiness, and then decreased and disappeared when spindles were present. They concluded that "SEMs seem to be associated with a process preceding or initiating drowsiness and subsiding as soon as drowsiness reaches a certain level, 20 or more seconds after its onset".

3.3.3 Performance fluctuations

The performance of subjects on tasks requiring continuous attention and their awareness of the external environment becomes intermittent during the transitional period between wakefulness and sleep (Makeig *et al*., 2000). Makeig (1995) observed that an increase of EEG power in the low theta range (2–5 Hz) corresponded with degraded visuospatial performance during a 2-D compensatory tracking task (Makeig and Jolley, 1996) and an auditory detection task (Makeig and Jung, 1996). In addition, Makeig and Jung (1996) found that these intermittent performance decrements occurred with a cycle of 15–20 s during their auditory detection task. The level of performance in behavioural tasks was negatively correlated with SEM activity during the transitional period (Torsvall and Åkerstedt, 1988).

3.3.4 Transition classification

The wake-sleep transition spans the standard R&K scoring stages, W (wake), stage 1 sleep, and stage 2 sleep. The key characteristics of these stages are:

> *Stage W*: Alpha activity and/or low voltage, mixed frequency EEG; blinks and rapid, darting eye movements. Relatively high EMG.

Stage 1: Low voltage, mixed frequency EEG, predominance of 2–7 Hz activity; less than 50% alpha activity in an epoch. No spindles, K-complexes, or REMs; slow rolling eye movements and vertex waves often present.

Stage 2: Sleep spindles, K-complexes; low voltage, mixed frequency EEG; may contain up to 20% delta waves (delta >75 μV and <2 Hz).

Although sufficient for staging 7–8 hours of nocturnal sleep, the R&K criteria are unsatisfactory for identifying the rapid transients that occur during the wake-sleep transition (Badia *et al.*, 1994). This is due to the relatively long epoch length of 30 s and the "averaging" that occurs as a result of the rater having to assign a single state label to a 30-s epoch. The dynamics of the EEG shifts rapidly during the sleep onset period and these important changes are lost during the rating process.

In order to classify the wake-sleep transition more accurately, Hori *et al*. (1994) subdivided the R&K stages W, 1, and 2 into nine EEG-based stages traversed sequentially as a person goes from wakefulness to unequivocal sleep. Typical EEG tracings for Hori's 9 stage scale are shown in Figure 3-2. The Hori stages in this text will be referred to with the prefix "H" to avoid confusion with the standard R&K stages. Stage W has been subdivided into 2 stages (H1 and H2), stage 1 into 6 sub-stages (stages H3–H8) and stage 2 sleep corresponds to stage H9 in the Hori scale. Hori *et al*. (1994) described their 9 stage partitioning as follows:

Stage H1: Alpha wave train: epoch composed of a train of alpha activity with a minimum amplitude of 20 μV.

Stage H2: Alpha wave intermittent (A): epoch composed of a train of more than 50% of alpha activity with a minimum amplitude of 20 μV.

Stage H3: Alpha wave intermittent (B): epoch contained less than 50% of alpha activity with an amplitude of 20 μV.

Stage H4: EEG flattening: epoch composed of suppressed waves less than 20 μV.

Stage H5: Ripples: epoch composed of low-voltage theta wave (20–50 μV) burst suppression.

Stage H6: Vertex sharp wave solitary: epoch contained one well-defined vertex sharp wave.

Stage H7: Vertex sharp wave train or bursts: epoch contained at least two well-defined vertex sharp waves.

Stage H8: Vertex sharp wave and incomplete spindles: epoch contained at least one well-defined vertex sharp wave and one incomplete spindle: duration <0.5 s, amplitude >10 μV, <20 μV).

Stage H9: Spindles: epoch contained at least one well-defined spindle of at least 0.5 s duration and 20 μV in amplitude.

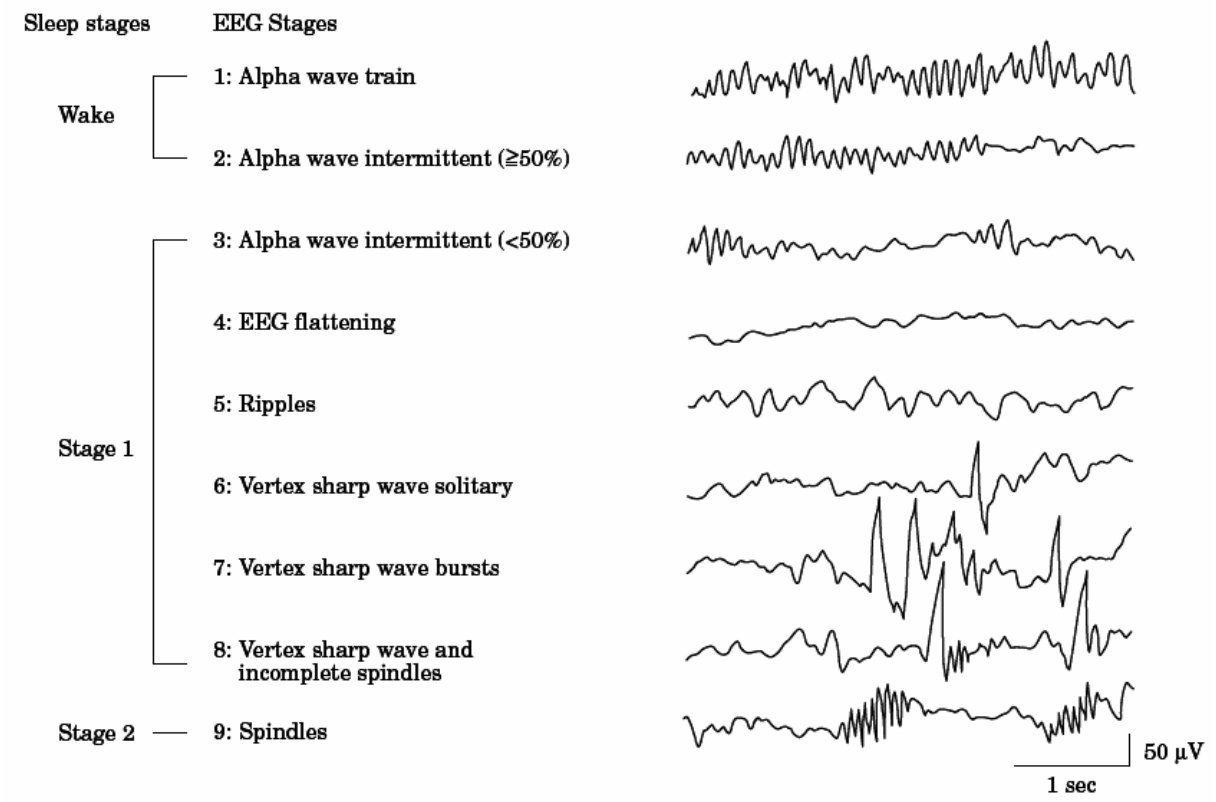

Figure 3-2 EEG recordings illustrating the 9 EEG stages during the wake-sleep transition according to the Hori *et al.* classification scale, and its correspondence with standard sleep stages (Hori *et al.*, 1994).

In contrast to the 20–30 s epoch length used in the R&K rating methodology, Hori used a 5-s epoch length to enable the capture of the rapid transients that occur during this period by preventing averaging over longer epochs.

Hori *et al.* (1994) found good behavioural, subjective, and physiological evidence for creating a 9-stage scale. They found a perfect ordinal relationship between RT and the 9 stages with the mean reaction time increasing from stages H1 through to H9 (711.1 ± 54.3 ms *vs*. 1936.9 ± 135.3 ms). Similarly, 82.5% of their subjects reported being "awake" during stage H1. This value decreased linearly to 26.2% in stage H9. As can be seen above, Hori's research provides strong evidence to suggest that subjects are not completely unresponsive during standard stage 2 sleep as indicated by the prolonged, but not absent, reaction times. Hori's work suggests that unequivocal sleep (i.e., zero responsiveness) begins post stage H9. Ogilvie (2001) stated that the partitioning of the wake-sleep transition using EEG features proposed by Hori *et al*. (1994) has contributed greatly to the understanding of the transitional period between wakefulness and sleep.

CHAPTER 4

Drowsiness Estimation and Lapse Detection: A Review

Chapter 2 presented a review of various physiological measures used to investigate arousal and attention, and Chapter 3 provided a review of the wake-sleep continuum. This chapter presents a review of knowledge and research in the areas of drowsiness estimation and lapse detection.

Researchers have been analysing the wake-sleep transition for over 6 decades. However, despite the growing body of research, the lapse/drowsiness research area is riddled with different interpretations of concepts and terms; it still lacks a set of accepted standard definitions. Numerous terms have been used to describe essentially similar phenomena. Reasons for this include a wide range of methods and measures and a lack of adequate documentation in some studies.

It is generally accepted that the EEG during drowsiness and sleep has more complex and variable patterns than the EEG during alertness (Maulsby *et al.*, 1968). A given individual can exhibit a variety of different patterns each time they traverse the wake-sleep continuum.

The first section of this chapter provides an overview of the most common interpretations and definitions of the key terms used in this research area, to help the reader grasp the remainder of this review chapter. Following this, a brief overview on automated sleep-staging is provided, and a review of approaches used to estimate drowsiness and detect lapses, such as EEG, eye-movement, head position, and video-based systems, is presented.

The chapter concludes with a statement of the research hypotheses underlying the research presented in the thesis.

4.1 Terminology

Overall, there is a lack of standard definitions for the key concepts in this area of research. Experienced researchers use certain terms interchangeably, often causing confusion for persons relatively new to the field. This section clarifies the definitions used throughout the thesis.

4.1.1 Alertness

There is definitional overlap in the terms *alertness* and arousal but the former can be distinguished by its focus on cognitive processing (Oken *et al.*, 2006). Freund *et al*. (1995) define loss of alertness as "an operator's internal state, during which time the processing of input information fails to reach the conscious cognitive threshold necessary for the operator to provide an appropriate response". They suggest that this phenomenon results from a complex interaction between factors such as boredom, physical and mental workloads, environmental stressors such as temperature, vibration, and glare, sleep quality and quantity, and circadian effects. The terms "loss of alertness", "loss of vigilance", "fatigue", "lapses/blocks", and "drowsiness" are often used interchangeably in the literature (Bittner *et al.*, 2000; Freund *et al.*, 1995).

4.1.2 Arousal

Arousal refers to a physiological state involving the level of activation of the nervous system which, from the perspective of human cognitive function means cortical activation (Segalowitz *et al.*, 1994). The terms 'cortical activity' and 'arousal' are used synonymously in the literature. According to Kahneman's classic capacity model (see Figure 4-1) of the attention-arousal system, arousal is the level of cortical activation driven from the reticular activating system and is considered a "low level" process (Kahnemann, 1973). The level of arousal determines the *attentional capacity* (attentional resources) available to a person (i.e., the person's alertness). By this definition, a low level of arousal leads to low attentional capacity and vice versa. However, very high levels of arousal can lead to attention being focused beyond the point of efficient task solving (Kahnemann, 1973) or to increased distractibility (Naatanen, 1975; Tecce *et al.*, 1976). Motivational factors can also influence the arousal level of a person (Hockey, 1984). For example, an interesting task results in increased effort from a subject, which leads to an increased arousal level. Similarly, a boring or monotonous task can result in a drop in effort and subsequent drop in arousal level. The

definition suggested by Oken *et al*. (2006) for arousal ("non-specific activation of the cerebral cortex in relation to sleep-wake states") is used in this thesis.

4.1.3 Attention

Attention refers both to the global capacity to focus mental processes on some experience (either endogenous or exogenous) and to the process of allocating this information-processing capacity (Segalowitz *et al.*, 1994). It is presumed that the nervous system has limited capacity and a mode of allocating the available attentional resources. Sustained attention requires motivation and cognitive processing and is "thought to be modulated by metabolic systems and substrates such as thyroid, glucose, oxygen and electrolytes" (Oken *et al*., 2006).

4.1.4 Drowsiness

Drowsiness is the transitory state between fully awake and sleep, and is 'stage 1 sleep' according to R&K criteria (Rechtschaffen and Kales, 1968). Several EEG and eye movement features can be used to identify drowsiness in a subject. Since changes in eye movements during drowsiness are more consistent than the EEG changes, drowsiness has been defined in terms of EOG changes by some researchers (Santamaria and Chiappa, 1987). Santamaria and Chiappa (1987) identified key EOG changes that occur during drowsiness to be "the disappearance of large blinks and fast eye movements, and the appearance of three types of eye movements (small-fast-irregular, small-fast-rhythmic, and slow)". Groups have reported an increase in delta/theta activity with increased drowsiness (Belyavin and Wright, 1987; Maulsby *et al*., 1968), as well as decreases in alpha and beta. Badia *et al*. (1994) reported that the largest decreases in relative EEG spectral power in 1-Hz frequency bins during drowsiness occurred around the alpha band (9, 10, and 11 Hz bins) and the greatest power increases were observed in the 3 and 4 Hz bins. Drowsy individuals performing an active task, such as driving, often cycle rapidly between periods of wake and sleep, as exhibited by cyclical variation in both EEG power spectra and task performance measures (Makeig *et al*., 2000).

4.1.5 Fatigue and sleepiness

The term *fatigue* is usually used to describe "physical and physiological fatigue from muscular exertion" (Freund *et al*., 1995) but has also been used to describe changes in both mental and physical processes after a sustained period of concentration on a task. The term has also been used interchangeably with "sleepiness" (Lal and Craig, 2002). Shen *et al*. (2006) stated that the "distinction between fatigue and sleepiness remains somewhat obscure",

"…neither sleepiness, nor fatigue (are) unitary phenomena, each in themselves being complex, heterogeneous phenomena", and that sleepiness and fatigue often tend to coexist as a result of sleep deprivation.

4.1.6 Microsleep

The term *microsleep* is often used to describe a brief period of sleep identified by an EEG dominated by theta activity (4–7 Hz) and an absence of alpha activity (8–12 Hz) (Harrison and Horne, 1996). The duration of a microsleep is stated as being from a minimum of 1–15 s to a maximum of 14–30 s (Harrison and Horne, 1996; Hemmeter *et al.*, 1998; Priest *et al.*, 2001; Tirunahari *et al.*, 2003; Valley and Broughton, 1983). There is a marked reduction in behavioural responsiveness during these periods. These events "show EEG characteristics similar to early stage 1 sleep" (R&K criteria) (Harrison and Horne, 1996) or stage 2 sleep (Valley and Broughton, 1983). Slow eye movements may be present during a microsleep event but the presence of alpha waves signifies wakefulness and rules out the possibility of a microsleep (Priest *et al.*, 2001). It is important to note that EEG-defined microsleeps are not always associated with reduced behavioural responsiveness. The term 'microsleep' will be used in the remainder of the thesis to refer to EEG-defined microsleeps.

4.1.7 Behavioural microsleep

Behavioural microsleeps (BMs) are characterized by eyelid closure, head-nodding, and markedly reduced or absent task responsiveness. They are frequently unaccompanied by EEG indications of sleep (Ogilvie, 2001).

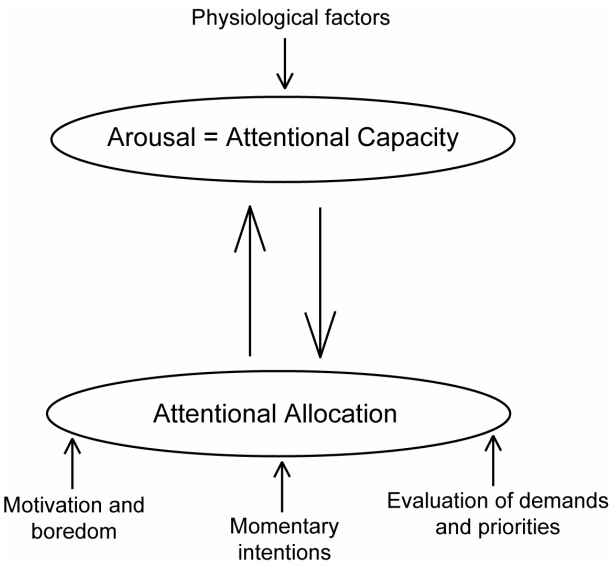

Figure 4-1 Schematic of the attention-arousal model outlined by Kahnemann (1973).

4.1.8 Vigilance

In the scientific community, the term *vigilance* has been used to mean sustained attention (Davies and Parasuraman, 1982; Oken *et al.*, 2006; Parasuraman *et al.*, 1998; Stroh, 1971). Specifically, 'vigilance decrement' is used to refer to the deterioration of attention and the consequent decline in task performance over extended periods of time (Mackworth, 1964). Animal behaviour scientists and psychiatric clinicians have used the term to mean attention to potential threats of danger, and also to refer to the level of cortical arousal, irrespective of the behavioural responsiveness and cognition levels (Oken *et al.*, 2006). In this thesis, the first definition of the term (i.e., sustained attention) is implied. Oken *et al*. (2006) have pointed out that despite being conceptually different from arousal, most studies on vigilance have examined variations in arousal level through the use of subjects who were either sleep-deprived, had sleep disorders, or were on sedative medications. In the same paper, Oken *et al*. also suggest that this problem is compounded due to "relative sleep deprivation" being common in the "overtly healthy population" as shown in work by Bonnet and Arand (1995) and Levine *et al*. (1988).

4.1.9 Lapses

In this thesis, the term *lapse* is an umbrella term which includes, unless otherwise stated, BMs, lapses of sustained/focused attention while in a state of alert arousal, and lapses of diverted attention. A lapse is considered to involve a complete temporary loss of performance/responsiveness.

4.2 Automated sleep-staging

Patients with sleep disorders are often required to undergo polysomnographic measurements including EEG, EOG, EMG, ECG, respiratory, and other signals as part of the diagnostic process. Traditionally, the sleep study is manually scored using the R&K criteria (Rechtschaffen and Kales, 1968) to gain insight into the sleep architecture throughout the night. However, the manual rating procedure is subjective, tedious, and time-consuming. A literature review by Agarwal and Gotman (2001) found that manual sleep scoring has relatively low inter-scorer agreement (67–91%). For these reasons, researchers have attempted to automate the sleep-stage scoring process.

According to Agarwal and Gotman's (2001) review into automatic sleep-scoring, approaches have included the use of period analysis (Itil *et al.*, 1969), EEG spectra with multiple discriminant analysis (Larsen and Walter, 1970), a combination of analog and digital techniques (Smith and Karacan, 1971), pattern recognition (Martin *et al.*, 1972), wave

detection with a Bayesian approach (Stanus *et al.*, 1987), interval histogram methods (Kuwahara *et al.*, 1988), expert systems (Ray *et al.*, 1986), and neural networks (Schaltenbrand, 1996). There is also segmentation and self-organization (Agarwal and Gotman, 2001), multi-fractal singularity spectra of the EEG and its Holder exponent (Ma *et al.*, 2005), and a self-organizing feature map with a fuzzy reasoning-based classifier (Tian and Liu, 2005). Compared to visual scoring agreements, the accuracy of automated techniques range from 75–85% (Scher, 2004). There is evidence that at least some automated sleep-staging systems available in the market are not sufficiently accurate to analyse sleep studies according to the R&K classification system. For example, Caffarel *et al*. (2006) reported an overall agreement (in terms of median kappa) of 0.331 between a commercial automatic sleep scoring system and a human rater, scoring 28 sleep studies into 'wake', 'light-sleep', 'deep-sleep', and 'REM' stages. In comparison, the agreement between 2 human scorers was higher (kappa = 0.641) when scoring the same dataset. Human scoring appears to be the generally accepted gold standard for rating sleep studies. However, the adoption of this standard is questionable given the low inter-scorer agreement reported in some reviews (Agarwal and Gotman, 2001).

It has also been reported that individuals have an EEG *fingerprint* during sleep and that this feature can be used to distinguish them during non-REM sleep (De Gennaro *et al.*, 2005). The EEG fingerprint is reported to be related to the topographic distribution in EEG power along the anterior-posterior cortical axis. It has been suggested that these EEG variances may be related to individual differences in "genetically determined functional brain anatomy" (De Gennaro *et al.*, 2005). It is possible that this may, at least, partly account for the lack of a highly accurate (say >95% agreement with a panel of human experts) sleep staging system to date.

4.3 Approaches to drowsiness estimation and lapse detection

The focus of this research project is on the transitional phase between the awake and sleep states. In this thesis, drowsiness is defined as the transitional period between awake and sleep (§ 4.1.4) and can be thought of as part of the arousal continuum. Conversely, reference to 'lapses' in the thesis imply discrete events where a temporary complete loss of performance occurs (e.g., auditory, visual, etc.) without the implication of a particular state of arousal. That is, the occurrence of a lapse does not necessarily imply that the subject is drowsy, and vice versa.

Most research into the awake-sleep transition has involved correlating the electrophysiological signals (EEG and EOG) with the performance level of a subject to identify the best indicators of a subject's level of vigilance. However, research groups have provided conflicting evidence regarding the relationship between physiological events and performance. These disparities appear due to the range of methods and measures used by researchers, as well as the fact that their studies have often not been well documented. It is generally accepted that drowsiness and sleep have more complex and variable patterns than the alert EEG pattern (Santamaria and Chiappa, 1987). While several EEG indicators are clearly correlated with reduced levels of vigilance and performance, none have proven to be highly reliable, at least over short periods of time.

There are no completely reliable and generally accepted automatic methods for analysing the EEG during studies of vigilance. Alpha activity, generally considered a reliable indicator of drowsiness, is of little use in performance estimation in subjects whose eyes are open (Jung *et al.*, 1997). Eye movements have provided a better indicator of a reduced level of alertness. SEMs have been observed during drowsiness and sleep onset and they became larger and more sinusoidal when coincident with EEG slowing due to drowsiness (Santamaria and Chiappa, 1987). SEMs are generally considered to occur at some stage during drowsiness in all subjects. Similarly, the disappearance of blinks and mini-blinks are a reliable sign of drowsiness and has been found to be very useful in scoring drowsiness. It is clear that incorporating eye movement activity into a drowsiness detection system will provide a superior estimation of alertness than a system based on EEG data alone.

Despite considerable research since the 1940s into the awake-sleep transition, no commercial device or technique capable of providing a highly reliable indicator of all levels and stages of the awake-sleep transition with high temporal resolution has been developed. There is contradictory evidence on the best EEG indicators of drowsiness. Hence, there is a need to identify features in the EEG which are highly reliable in indicating and predicting drowsiness and lapses in a subject. Current technologies also appear to have difficulty coping with the inter-individual variability observed in the drowsiness EEG. For example, performance of techniques such as power spectral analysis improve when customized to each individual user before being used (Belyavin and Wright, 1987; Jung *et al.*, 1997; Lin *et al.*, 2006; Lin *et al.*, 2005a; Lin *et al.*, 2005b; Makeig *et al.*, 1996). This is time consuming and reduces its usability. Even then, purely EEG-based techniques such as power spectral analysis have been unable to produce highly reliable measures. Research is also lacking in the area of accurately detecting and predicting lapses in performance from the EEG and/or EOG. The accuracy of current drowsiness detection technologies is insufficient to warrant large-scale application in the transport industry. They also appear unable to predict the onset of lapses.

Researchers have attempted to incorporate head position and eye movement cues in the form of video-based real-time drowsiness monitoring devices, targeted predominantly for the transport sector (Bergasa *et al.*, 2006; Ji and Yang, 2002).

Sections 4.4 and 4.5 summarize the approaches that have been used in the estimation of drowsiness, and detection of lapses, respectively.

4.4 Drowsiness estimation

This section provides a summary of the key EEG-based, eye-closure-based, head-position-based, and video-based approaches used to date in the area of drowsiness estimation. Studies which utilize eye-movements and eye-closure as a secondary measure to the EEG are also included in § 4.4.1.

4.4.1 EEG-based approaches

Matousek and Petersen (1983) were one of the first to develop a method for the automatic evaluation of the vigilance level using the EEG frequency pattern. Their study included 477 epochs of 5-s duration from routine EEGs of 41 subjects. Epochs considered to represent either fully alert or stage 1 sleep were selected by visual inspection. The selected epochs were marked either '0' (alert) or '1' (sleep). The test material consisted of 265 epochs belonging to another set of EEGs. Twenty two variables were calculated based on the EEG spectra using 8 EEG channels (F7-T3, F8-T4, C3-C0, C4-C0, T3-T5, T4-T6, P3-O1 and P4-O2). The spectra were divided into 6 bands (delta, theta, alpha 1, alpha 2, beta 1, and beta 2) with limits 1.5, 3.5, 7.5, 9.5, 12.5, 17.5 and 25 Hz. The mean values for all 22 variables (only considering 'alert' epochs) were calculated. All values belonging to both 'alert' and 'sleep' epochs were then normalized against the corresponding 22 mean values. The EEG data was related to the assigned level of vigilance using multiple regression.

Both normalized and non-normalized EEG variables had significant correlations with the vigilance level (i.e., either 'alert' or 'sleep'). The correlation was stronger for normalized values. Beta activity contributed significantly to the final vigilance estimate. The vigilance estimate was obtained by summing all the products between the EEG variables and a set of weighting factors, offset by a constant. Thus

$$\textit{Calculated vigilance level} = A(0) + A(1)x(1) + \ldots + A(22)x(22), \qquad (4\text{–}1)$$

where $A(0)$ is a constant, $A(1)$-$A(22)$ are the weight coefficients, and $x(1)$-$x(22)$ are the EEG variables.

Matousek and Petersen (1983) found that automatic estimates were in good agreement with the individual visual findings in the test material with a correlation coefficient of 0.86. However, they state that "all these computerised methods were inferior to visual assessment in the presentation of the results". Trials using more categories than '0' and '1' did not improve the performance. The authors also did not find it necessary to supplement the EEG with other cues such as SEMs. The correlation between the automatic and visual findings of alertness was 0.89 in the training material and 0.86 in the test material.

Belyavin and Wright (1987) observed changes in the EEG as vigilance deteriorated during a long period of repeated monitoring tasks, which included daytime and overnight work. The study included vigilance and discrimination tasks. EEG from P3-O1 and P4-O2 with nasion as ground, EMG from neck muscles, and EOG from above and below the right outer canthus were recorded. The EEG was recorded on a 16-channel polygraph. Power spectra over 1–30 Hz were calculated using the FFT and were averaged over 3 s. The averaged spectra were divided into bands delta (1–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta1 (14–21 Hz), and beta2 (22–30 Hz). Statistical analysis was carried out on the data from channel P3-O1 since preliminary evaluation had shown a close relationship between vigilance performance and this particular channel. Changes in the EEG, as the proportion of stimuli missed increased, were investigated by calculating canonical variates. (In this type of analysis, linear combinations of EEG variables are derived to maximize the difference between groups while minimizing the within-group variance. This can be used to determine whether the variance in the EEG between groups arises from one source or whether there is more than one independent source). Canonical variates showed the existence of a complex relationship between the EEG and vigilance. Increased delta and theta activity, as well as decreased beta1, was associated with worsening performance levels. Theta and beta1 contributed substantially to the regression across subjects for a simple vigilance task. Performance was more closely related to the EEG in the vigilance task than in the discrimination task. Canonical variate analysis indicated that two directions were involved in the relationship between EEG and vigilance in both tasks. The presence of a second direction indicated non-linearity in the relationship between the changes in the EEG and performance. It was concluded that "beta activity is the most useful discriminator of worsening vigilance common to both tasks" (Belyavin and Wright, 1987). The authors also suggested that pooling subject data may result in the loss of valuable information by destroying subtle differences in the individual EEGs.

Makeig and Inlow (1993) conducted a study using 13 male subjects on a passive sonar auditory target detection task (eyes closed). The subjects were asked to press a buzzer as soon as a target noise burst was detected and to ignore the probe tones. EEG signals were recorded from 13 scalp locations using the 10-20 system, and EOG was recorded using

periocular electrodes. The local error rate was calculated for the analysis of the vigilance test results. This provided a continuous estimate of performance at regularly spaced time intervals. The local error rate was derived by computing the fraction of undetected targets within a time window with a constant width of 32.8 s, which was advanced through the data in steps of 4.64 s. Coherence analysis was used to detect linear relationships between fluctuations in performance and changes in the EEG power spectrum. It was used to measure the degree to which slow changes in the EEG power at one site accompany slow changes in the local error rate. A non-parametric Monte Carlo procedure was used to assess the significance of the results. They observed an increase in the mean error rate after the first 2–3 min. However, numerous unpredictable fluctuations in performance resulted within individual runs. A reduction of 10 Hz and the emergence of 4 Hz EEG activity were observed when the error rate increased. It was also found that three narrow EEG spectral bands, with maxima near 8, 10, and 13 Hz, dominated the coherence plane at channel Cz (central). Fluctuations in EEG power below 6 Hz were in phase with error rate, whilst fluctuations above 7 Hz were nearly 180° out of phase. There was no appreciable time lag between changes in error rate and EEG power. Analysis showed that changes in performance are linearly related to specific patterns of changes in the EEG power spectrum at all EEG recording sites and over a wide range of performance cycle lengths. It was also found that this relationship is relatively variable but stable within subjects. The analysis also showed the limitations of using *a priori* frequency band averaging to study changes in the EEG spectrum. In summary, Makeig and Inlow (1993) found that changes in the EEG power spectrum constantly accompany slow and irregular fluctuations in arousal and cognitive state, which may last for a period under a minute or for several minutes. They also found that the exact pattern of correspondence between EEG and behaviour differs between individuals. However, they report that the relationship remained stable within individuals across sessions and, that individualized models can be used to predict the time course of performance during drowsiness.

Makeig and Jung (1995) found that a single principal component of EEG spectral variance is linearly related to minute-scale changes in detection performance in a group of 15 subjects engaged in an eyes-open dual-task (auditory and visual) paradigm. All subjects performed 5 sessions of 30 min each. EEG was recorded at Cz and Pz/Oz referenced to the right mastoid at a sampling rate of 312.5 Hz (pass-band 0.1–100 Hz). Normalized log spectra were calculated at Cz and Pz/Oz for 40 frequency bins between 0.61 and 24.4 Hz, and were smoothed using a 95-s exponential window. The discontinuous, irregularly-spaced auditory detection performance index was also smoothed using the 95-s exponential window to obtain a continuous local error rate. Two sessions each from 10 (of the 15) subjects were selected for further analysis. A "correlation spectrum" was calculated to find the correlation between smoothed log spectral power and local error rate for all 40 frequency bins for each

of the 20 sessions. The mean correlation spectrum was calculated by averaging over the 20 sessions. A 'grand correlation spectrum' was also calculated using data from all 20 sessions with the normalized data described above. They found that mean and grand correlation spectra were nearly identical (r = 0.995, RMS difference = 0.023). The correlations between EEG power and performance were significant in four frequency bands: at 4–5 Hz (theta) in both Cz and Pz/Oz, at 10–11 Hz (alpha) at Pz/Oz, at 14–15 Hz (sigma) at Cz and above 15 Hz (beta) at Pz/Oz. The EEG frequencies at which log power was correlated to auditory detection performance was similar for most subjects.

Makeig and Jung (1996) observed that mean activity levels in the delta (<4 Hz) and theta (4–6 Hz) bands and "sleep spindle frequency (14 Hz)" were higher during periods of lowered performance in an auditory detection task, when compared to alert performance, in a group of 15 young adults (non-sleep deprived) in the study mentioned in the previous paragraph (Makeig and Jung, 1995). The tests were conducted in a small, warm, and dimly-lit experimental chamber. They also observed that the 4–6 Hz theta activity increased approximately 10 s before an undetected auditory target ("lapse"). A decrease in gamma band activity above 35 Hz was also observed 10 s prior to the missed targets. They found that the theta and gamma frequency EEG power decreased to baseline about 9 s after stimulus presentation. The authors also noted that alpha activity was not phase locked to the subjects' auditory performance fluctuations. A study by Cajochen *et al*. (1999) confirmed Makeig's and Jung's findings and observed increased delta and theta band activity correlated with a reduction in PVT performance (visual vigilance task) in a study involving 10 males undergoing a constant routine protocol with 32 hours of wakefulness. They also reported a poor correlation between alpha activity in both frontal and occipital derivation and PVT performance. Makeig *et al*. (2000) also found a trend similar to their auditory target detection study (between EEG changes and fluctuations in performance) in a continuous, visuospatial, tracking study involving sleep-deprived subjects. They suggest that the similarity observed between EEG changes and performance in auditory and visual tasks implies that they may be common to any sensory-motor task. For both modes of tasks, the most prominent feature was an increase in power in the 2–5 Hz band with degraded performance. This occurred regardless of whether the data was averaged over seconds or minutes. The authors report the occurrence of "wake-like" and "sleep-like" performance cycles during periods of drowsiness with a cycle period of around 18 s.

Jung *et al*. (1997) conducted an experiment using 15 subjects on a simulation of an auditory and visual sonar target detection task. Each person participated in three or more simulated work sessions, with each session lasting 28 min. EEG data was recorded at a sampling rate of 312.5 Hz from two midline sites, one central (Cz), and the other mid-way between parietal and occipital sites (Pz/Oz). Moving-averaged spectral analysis of the EEG data was then

performed using a 256-point Hanning window with 50% overlap. Windowed 256-point epochs were extended to 512 points by zero padding, followed by Fourier transformation of the resulting 512 points. Median filtering using a moving 5-s window was used to further minimize the presence of artifacts in the EEG records. The local error rate within a moving time window was calculated. Each error rate time series consisted of 1024 points at 1.64-s intervals, computed using a causal 93.4 s (57 epoch) exponential window whose gain varied from 1.0 at the leading edge to 0.1 at the trailing edge. They found strong evidence for a monotonic relationship between fluctuations in EEG power and performance in two relatively narrow bands, near 3.7 Hz and 14.7 Hz. The EEG power increased in both bands as the local error rate increased. They also suggest that alpha frequencies (8–12 Hz) may be of relatively little use for error rate estimation in subjects whose eyes are open. The mean correlation between performance and EEG power was positive at both sites (Cz and Pz/Oz) around 4 Hz, and at Cz around 14 Hz. At high error rates, a modest negative correlation also appeared near 10 Hz. They state that standard frequency bands cannot be used to accurately predict individual changes in alertness and performance because information about alertness may be distributed over the entire EEG spectrum. The authors have investigated the feasibility of practical alertness monitoring by combining power spectrum estimation, PCA, and artificial neural networks (ANNs). They found that the most effective performance estimate was obtained by using 4 principal components. The non-linear adaptability of multi-layer perceptrons improved estimation performance over linear regression, reducing the RMS error on the testing data across subjects. The best performance (in terms of mean RMS estimation error for 20 sessions from ten subjects) obtained was 0.156 ± 0.048 for a neural network with 1 hidden layer with 3 units. Their results showed that accurate, individualized alertness estimation, using neural networks applied to EEG spectral data, was possible. Jung *et al*.'s results in applying neural network estimation to the full EEG spectrum compared favourably with previous research, which applied a linear regression model to pre-selected frequencies.

A study by Dinges (1998) (see § 4.4.2 for methodological details), assessed the validity of the system developed by Makeig and Inlow (1993) to detect lapses in visual attention (measured via a PVT task) from the EEG. They reported a mean between-session correlation of 0.62 ± 0.16 (N=4) and a minute-to-minute correlation of 0.46 ± 0.10 (N=4) with lapse frequency.

Huang *et al*. (2001) developed a procedure to obtain relatively efficient and reliable EEG features, with the least redundancy, for predicting alertness levels. Thirteen subjects participated in an auditory detection task and seven in a visual task. In the auditory task, subjects were asked to press a button whenever a specific tone was heard. Similarly, subjects were asked to press a button when a red LED flashed during the visual detection task. Both

stimuli had a random inter-stimulus interval between 3–8 s, with an average of 5.5 s. A total of 700 stimuli were presented in each 45-min experimental session. One bipolar EEG was recorded using electrodes placed at C3(+) and C4(-) with Cz grounded. Chin EMG was employed to verify movement artifacts in the EEG. The behavioural alertness level was described by the correct rate (CR), which was defined as the ratio of the number of correct responses to the number of auditory or visual stimuli within a 32-s rectangular moving window. They extracted EEG features in three dimensions (time, frequency, and statistics). In the time domain, the statistical mean value, sum of the absolute amplitudes, sum of the squared amplitudes, standard deviation, 3rd moment about the mean amplitude, skewness coefficient, 4th moment about the mean amplitude, and kurtosis coefficient, were calculated. Sixteen other features were extracted from the EEG amplitude spectrum using the FFT with a Hanning window: sum of the full (0–32 Hz, $|A|$), β(13–32 Hz, $|\beta|$), α(8–13 Hz, $|\alpha|$), θ(4–8 Hz, $|\theta|$), and δ(0–4 Hz, $|\delta|$) bands; the relative spectral amplitudes $\beta(\beta\%)$, $\alpha(\alpha\%)$, $\theta(\theta\%)$, and $\delta(\delta\%)$ bands; the mean frequencies in the full (MF), $\beta(F\beta)$, $\alpha(F\alpha)$, $\theta(F\theta)$ and $\delta(F\delta)$ bands; the sum of the squared spectral amplitudes, and the standard deviation of the frequency spectrum.

For the auditory detection task, the most suitable features were α%, θ% and MF. The mean squared correlation coefficients (r^2) between CR and selected EEG features, and the squared multiple correlation coefficients (R^2) between CR and a subset (α%, θ% and MF) were calculated. Mean R^2 values were 69% for the 32-s sub-window and 75% for the 4-s sub-window. It was noted that "the application of averaging sub-window analysis within a moving time window seemed to be a valuable procedure for reducing the disturbing effect of movement artifacts on the EEG". For the visual task, the single best predictive feature was $F\beta$. Mean R^2 values were 19% for the 32-s sub-window and 23% for the 4-s sub-window. However, the mean R^2 between CR and $F\beta$ was relatively small compared to that between CR and the 24 EEG features (19% *vs*. 44% in the 32-s sub-window and 23% *vs*. 54% for the 4-s sub-window).

Vuckovic *et al*. (2002) applied 3 ANNs – linear network trained by Widrow-Hoff rule, feed-forward network trained with the Levenberg-Marquardt learning rule, and learning vector quantization (LVQ) – to EEG data collected from 17 normal, non-sleep-deprived subjects during normal working hours and assessed its classification performance. The subjects were lying in a dark room with their eyes closed, and a neurologist in the room ensured that they did not fall asleep further than stage 1 sleep. The experimenters wanted to conduct the experiment such that there was no "interference from transient change in alertness induced by cognitive tasks" and, hence, there was no behavioural task associated with the study. The duration of the recording sessions was 30 min and electrodes were placed at 14 scalp locations (F7, F8, T3, T4, T5, T6, F3, F4, C3, C4, P3, P4, O1, O2) according to the 10/20 system

using an average reference. Left and right EOG was also recorded and all data were sampled at 256 Hz. Two expert EEG raters independently identified clear 'alert' and 'drowsy' 1-s epochs using the EEG and EOG data. Sequences where the two raters disagreed were discarded. Moving cross spectral density (CSD) was calculated with signals lasting 1 epoch. Inter-hemispheric CSD was calculated from homologous contralateral electrodes and intra-hemispheric CSD was calculated from electrodes on the left hemisphere. In addition, the authors also introduced 3 consecutive CSD values to make a "time-shifted CSD" which yielded a total of 42 CSD parameters as input to all ANNs. The training set consisted of 20 epochs of drowsiness and 20 epochs of alertness from each subject. The validation set consisted of 40 epochs of alertness and 40 epochs of drowsiness from each subject. The authors found that the LVQ network had the best classification performance out of all 3 ANNs. When trained on 5 of the 17 subjects, and validated using the remaining 12 subjects, the LVQ ANN showed an agreement of 94 ± 2% between its output and the expert rating. Analysis showed that both intra and inter-hemispheric CSD, and treating the CSD as a time series had an important impact on the network's generalizability.

Lal and Craig (2002) reported increases in delta, theta, and alpha (to a lesser extent) power correlating with increasing levels of "fatigue" during a continuous and monotonous driving task. Fatigue was identified via visual cues such as changes in facial tone, blink rate, eye activity, and mannerisms such as nodding and yawning, and it can be assumed to be equivalent to drowsiness/sleepiness. The authors reported observing SEMs leading to partial or full eye closure together with features such as head droops or continuous nods; in effect implying that the subjects were having behavioural microsleeps. The authors also investigated the reproducibility of EEG magnitude changes in the delta, theta, alpha, and beta bands during fatigue (Lal and Craig, 2005) and between two transitional episodes to fatigue, and found a high reproducibility level in the delta and theta bands ($r > 0.95$). The authors used magnitude changes in the four EEG bands to discriminate between 3 levels of fatigue (early fatigue – stage between awake and slow-wave activity; extreme fatigue – early stage 1 sleep dominated by slow-wave activity; medium fatigue – characteristics of both early and extreme fatigue) (Lal *et al.*, 2003). The largest detection error between the EEG-magnitude-change-based algorithm and video rating was found during the classification of early to moderate levels of fatigue with an average detection error of around 10% (Lal *et al.*, 2003). The authors also reported that their study showed that although alpha activity increases during drowsiness, there was a more substantial change in magnitude in the delta and theta bands and that this was easier to detect.

Subasi (2005b) proposed a system based on decomposing the EEG into frequency sub-bands using the wavelet transform, and extracting a number of statistical features from the sub-bands to represent the distribution of the wavelet coefficients. These features were used to

train an error back-propagation ANN to recognize different levels of "alertness". The EEG was rated into "alert", "drowsy", and "sleep" states by two experts. Only sections of data agreed by both raters as being clearly alert, drowsy, or asleep were used for the study. Data from 30 non-sleep-deprived subjects were collected and the network was validated using data from 12 of the subjects who were not used to train the ANN. Classification accuracies of 95 ± 3%, 93 ± 4%, and 92 ± 5% were obtained for alert, drowsy, and sleep states, respectively.

Building on the earlier work of Jung *et al*. (1997), Lin *et al*. (2005a) confirmed a close relationship between "minute-scale" (i.e., 60-s) changes in driving performance and the EEG power spectrum. Ten subjects completed a driving task in a virtual-reality-based simulator on two separate sessions (1–7 days apart). Data from subjects who had two or more microsleeps in each of the two sessions were selected for further analysis (N=5). Each session lasted 45 min. The authors used the deviation between the center of the vehicle and center of the road as the performance index. This metric was smoothed using a causal 90-s square moving-averaged filter advanced at 2-s steps. EEG was collected from 33 channels at a sampling rate of 250 Hz. The EEG was first pre-processed using a low-pass filter (cut-off = 50 Hz) to remove mains and other high frequency noise. The correlation coefficients between the smoothed driving performance index and log power spectra of all EEG channels at each frequency band were calculated to form what the authors termed a "correlation spectrum". The log power spectra of the 2 EEG channels with the highest correlation (Cz and Pz) were selected and decomposed using PCA. The first 50 principal components resulting from PCA were used as the input vectors to a linear regression model (with a least-square-error cost function) to estimate the subject's driving performance. Each model was trained using data from one session of a subject and tested on the data from the 2nd session of the same subject. The relationship between log power spectra and driving performance was stable within individuals across sessions. However, there was variability between subjects. The mean correlation coefficient between actual driving performance and the model's estimate was 0.90 ± 0.03 (within-session) and 0.53 ± 0.12 (between-session).

Visiting the same data, Lin *et al*. (2005b) evaluated an alternative method based on ICA, power-spectrum analysis, correlation analysis, and a linear regression model to estimate driving performance. ICA was used to remove artifacts from the 33 EEG channels and derive 33 ICA components. Normalized moving-averaged sub-band log power spectra were calculated for all ICA components. The driving performance metric was smoothed using a causal 90-s square moving-averaged filter advanced at 2-s steps. Correlation coefficients between the smoothed performance index and power spectra of all ICA components at each frequency band were calculated to form a correlation spectrum. Five frequency bins from the 2 ICA components displaying the highest correlation coefficients were selected as the input features to train a separate linear regression model for each subject. Following this, the

ICA "unmixing" matrix from the training session was used to spatially filter the testing data of a given subject. Data from the 2nd session was used to test the model. The mean correlation between estimated and actual driving performance was 0.86 ± 0.07 (within-session testing) and 0.88 ± 0.05 (between-session testing). The authors also compared performance when the two best EEG channels of each subject were used to train the model (cf. ICA components). This yielded mean correlations of 0.85 (within-session testing) and 0.82 (between-session testing), indicating that the use of ICA components improves performance.

Lin *et al*. (2006) extended their approach still further by adding an adaptive feature selection method to their EEG-based performance estimation system using the same dataset as described earlier (Lin *et al*., 2005a; Lin *et al*., 2005b), they smoothed the driving performance index using a causal 90-s square moving-averaged filter advanced at 2-s steps. An adaptive feature selection mechanism, based on correlation coefficients between log band power of the drowsiness-related ICA components and the subject's performance index, was utilized. An ICA-mixture-based fuzzy neural network was used to estimate the performance of the subject. The mean correlations between the model's output and the performance index were 0.98 ± 0.01 (within-session) and 0.91 ± 0.03 (between-session).

Papadelis *et al*. (2007) developed and evaluated an EEG-based driver alertness estimation system using both linear and non-linear analysis methods. Twenty-one subjects (20 M, 1 F), sleep-deprived for 24 hours, performed an on-road driving task. The task consisted of driving approximately 200 km on a monotonous highway. It was carried out using a modified car with double support pedals which were accessible to the instructor as well as advanced driver assistance systems and sensors such as a Lane Detection System (LDS) (to estimate position and orientation of the test vehicle via identification of lane borders) and infra-red eye-lid sensors to detect eye blink duration and per minute averaged blink duration. The LDS had a time resolution of 1 s and the eye-lid sensors had a millisecond resolution. Data from the LDS system verified by the driving instructor were used to identify driver error events. The signals recorded included EEG (Fp1, Fp2, C3, C4, P3, P4, O1, and O2), horizontal and vertical EOG, ECG, and EMG. A 50-Hz hardware notch filter was used and all channels were sampled at 200 Hz. The EEG was band-pass filtered using a 2nd order Butterworth filter (BW: 0.5–40 Hz). ICA was used to remove artifacts from the data. For each EEG channel, the relative band ratio (RBR) – power ratio of a specific band to the total frequency power – for the delta, theta, alpha, beta, and gamma bands was calculated. Shannon Entropy, Kullback-Leibler Relative Entropy, and Approximate Entropy (ApEn) were also calculated for each channel. An epoch size of 1-s was used for the above calculations. The above statistics were then averaged over 5-min periods. A linear coherence measure and Cross-ApEn were used to quantify the synchrony among pair-wise

channels and were also averaged over 5-min periods. The authors found that "brief paroxysmal bursts of alpha activity and the increased synchrony among EEG channels are strongly linked with an upcoming driving error". The duration of these bursts was 1–2 s. The authors reported increases in RBR between the first and last 15-min epochs in the delta, beta, and gamma bands. Decreases in Shannon Entropy and Kullback-Leibler Relative Entropy were found between the first and last 15-min epochs. The spectral RBR of alpha activity increased about 1-min prior to driving errors and was dominant in the central and parietal channels. Coherence increased before driving errors in the alpha and gamma bands. The cross-ApEn was found to be a more sensitive measure of detecting "sleepiness onset" and that per minute averaged blink rate increased in the minute preceding a driving error.

The 'Drowsiness Detection Algorithm' by Consolidated Research Inc. (CRI) (Euclid, Ohio, USA) utilized EEG data from a single occipital site (O1 or O2) referred to A1/A2. A window size of 2.4 s and an overlap of 50% between successive windows were used. The algorithm's output magnitude was proportional to the subject's alertness level. The developers claim that the algorithm is sensitive to second-by-second changes in alertness level due to its fast update rate and, hence, can successfully detect microsleeps and microarousals. The CRI EEG algorithm relies exclusively on EEG measures to track the state of arousal of a subject, although it is not clear what features are used in their algorithm. The company states that their algorithm output does not give a direct measure of performance in terms of reaction times but, rather, is an indirect measure of task performance via measurement of the arousal level. This algorithm showed a mean between-session correlation of 0.53 ± 0.06 (N=4) and a minute-to-minute correlation of 0.29 ± 0.06 ($N = 4$) with lapse frequency in an evaluation of the performance of six drowsiness measurement algorithms (Dinges *et al.*, 1998).

4.4.2 Eye-closure and eye-movement-based systems

The Blinkometer by IM System Inc. (Baltimore, MD, USA) is another eye-movement based device claimed capable of detecting drowsiness/sleepiness. It was designed to operate in blinks per minute mode or blink-to-blink interval mode. In Dinges *et al.*'s (1998) study, the blinks/min mode was used as the metric, as it yielded information not provided by the five other technologies under evaluation. The Blinkometer consists of a sensor placed on the outer canthus of one eye connected to a portable recording device. Blink activity was detected when a piezoelectric film in the sensor was moved by eyelid activity, which caused a red LED in the device to flash. The Blinkometer required calibration to adjust the blink detection threshold for individual subjects. The Blinkometer algorithm is based on the assumption that the blink rate reduces as the level of drowsiness increases. The mean blink rate of the baseline session was calculated during the first 20-min PVT session when subjects were considered fully awake. In subsequent test sessions, the blink rate of each 1-min epoch

was calculated and a drowsiness score (range: 0–5) assigned to the epoch by comparing it to the baseline rate. If the mean blink rate during the epoch was less than 1 SD below the baseline mean, a drowsiness score of 0 was assigned to the epoch. Similarly, if the mean blink rate during the epoch was between 1 and 2 SD of the baseline mean, a drowsiness score of 1 was assigned, etc. The Blinkometer developers claim that their algorithm detects a decreased alertness level within 20–30 s. The evaluation by Dinges *et al*. (1998) showed that the device had a mean between-session correlation of $r = 0.57 \pm 0.11$ (N=6) and $r = 0.29 \pm 0.08$ with PVT lapse frequency. However, there has been no mention in the literature or elsewhere of the Blinkometer since Dinges *et al*. (1998).

Like the Blinkometer, the Alertness Monitor by MTI Research, Inc. (Westford, MA, USA) is based on eye blinks to detect and track changes in drowsiness. It uses optical electronics to calculate the ratio between the degree of eye-lid closure to eye-lid open and, from that, determine the level of alertness. It achieved this via a device resembling a modified pair of spectacles, which contained unobtrusive optical electronics mounted such that an infra-red beam emitted from it fell along the axis of the eye blink and the eye lash would break the beam during an eye blink. The emitter is located on the nose-piece of the spectacles and the sensor is located on one of the arms of the spectacle. The device requires calibration for each subject and provides a single metric of drowsiness using MTI's proprietary algorithm. The Alertness Monitor showed a mean between-session correlation of $r = 0.33 \pm 0.10$ (mean ± SE, N=14) and a minute-to-minute correlation of $r = 0.22 \pm 0.07$ with lapse frequency (Dinges *et al*., 1998). Like the Blinkometer, there is no evidence to suggest that MTI's Alertness Monitor evolved beyond the prototype stage.

Van Orden *et al*. (2000) used five concurrent eye activity measures, obtained automatically via an eye-tracking system (Applied Sciences Laboratory SU4000 eye-tracking system), to model "fatigue-related changes in performance during a visual compensatory tracking task". The measures were moving estimates of blink duration and frequency, fixation dwell time and frequency, and mean pupil diameter. These features were analysed using non-linear regression and ANN techniques. Nine participants completed two 53-min test sessions on separate days from which continuous video-based eye activity and tracking performance measures were obtained. The authors derived individual models using eye and performance data from one session and cross-validated using data from the second session. A general regression model, based only on fixation dwell time and frequency from both sessions of all participants, was used to estimate the tracking performance and produced a correlation of $r = 0.68$ between estimated and actual tracking performance. The authors also found that blink duration was highly correlated with tracking error for many participants but not as robust or reliable as fixation dwell time. Individualized ANN models derived from data from both sessions produced the best performance in terms of highest correlation between

eye-measure-based tracking estimates and tracking performance ($r = 0.82$). Presumably, drowsiness and/or lapses may have contributed to the "fatigue-related changes in performance" that Van Orden and colleagues observed in their study.

Grace *et al*. (1998) proposed a video-based drowsy-driver detection system for heavy vehicles using the PERCLOS measure (see § 4.5.2 for a description). The system consisted of two infra-red light sources at different frequencies and two charged-couple-device cameras situated at a 90-degree angle to each other. The system used retinal reflection differences caused by the two infra-red sources to calculate the percentage of eye closure. The system is also reported to cope adequately with users wearing spectacles.

The 'Copilot' (Grace, 2001) was proposed as a second generation system to the original system. The device contained an auditory advisory tone and visual gauge to indicate the level of drowsiness to the driver. Validation of the Copilot system was said to be in progress at the time of the publication but, to date, there have been no further publications on this device.

Eye-Com, Inc. (Rena, Nevada, USA) is developing a device that can be used to monitor a user's eyes and detect signs of drowsiness. The device consists of a wearable, non-invasive solder helmet and an eye-frame mounted eye-tracker. Two CMOS IR-sensitive micro-cameras (angled upwards under each eye) and a single or linear array of IR illuminating LEDs is attached to the eye-glass frame. Using this set-up, the system can track both IR-illuminated pupils under a variety of lighting conditions to measure over 20 oculometric parameters. These parameters include eye blink duration, frequency, and velocity, PERCLOS, saccadic eye movement velocity, pupil size, and pupil response latency to light flashes. Since development of Eye-Com's drowsiness detection system is still in progress, no information about its performance is available in the public domain.

The Optalert Drowsiness Measurement System (Sleep Diagnostics Pty Ltd, Melbourne, Australia) uses a technique known as IR reflectance oculography to measure a subject's level of drowsiness. The hardware of the system included IR illuminators, an IR transducer, and a microprocessor and were mounted on a spectacle frame. Parameters measured by the system, such as relative velocity and duration of blinks and eye-lid closures were combined to obtain a drowsiness measure termed the Johns Drowsiness Scale score (Johns *et al.*, 2007). A study utilizing the Optalert system reported that the system predicted instances where a vehicle in a driving simulator was leaving the road 15-min in advance with a sensitivity and specificity of 83% and 61% respectively, in a group of 20 subjects who were sleep-deprived (Stephan *et al.*, 2006).

4.4.3 Head-position-based systems

The Proximity Array Sensing System (PASS) by Advanced Safety Concepts (Santa Fe, NM, USA) is based on the premise of increasing changes in head position with increasing levels of drowsiness. These are caused by a reduction in the tone of muscles in the neck and head causing the head to drop or roll on occasions.

The PASS system records the x, y, and z coordinates of the head using three electromagnetic fields and uses these to determine the level of drowsiness in the form of two output metrics termed ASC60 and ASC90. Background details of the two metrics are not available. ASC60 showed a mean between-session correlation of $r = 0.46 \pm 0.28$ (mean ± SE, N=5) and ASC90 a slightly higher mean correlation of $r = 0.52 \pm 0.15$ (mean ± SE, N=5) (Dinges *et al.*, 1998). The ASC60 and ASC90 indices showed a minute-to-minute correlation of 0.30 ± 0.12 and 0.26 ± 0.07, respectively.

4.4.4 Video-based systems

This section summarizes several approaches that incorporate features such as face orientation, head-movement, eye closure and/or eye movements to estimate drowsiness.

The 'Antisleep' system (http://www.smarteye.se/antisleep.html) measures a driver's head position, eye-gaze direction, and eye-lid opening using a single camera and two infra-red (IR) flash illuminators to detect their level of attention. However, although Antisleep is a commercial product, no performance either independently measured and reported, or claimed by the Antisleep developers was found.

An approach similar to the Antisleep system was proposed by Ji and Yang (2002), utilizing real-time eye tracking to measure eye-lid movements, gaze, and face orientation. Their system also remains under development, and there are no available validation study results in the public domain.

Bergasa *et al*. (2006) developed a prototype which used infra-red illumination and several image-processing algorithms to obtain an estimate of the "driver inattentive level (DIL)". Their system used parameters such as PERCLOS, eye-close duration, blink frequency, head-nodding frequency, face position, and fixed gaze, and combined them using a fuzzy system. System performance was analysed in terms of percentage agreement between the estimated DIL and manual human scoring on a frame-by-frame basis. An optimum system performance accuracy "close to 100%" was reported when tested on a sequence of simulated fatigue behaviours at night for a group of users not wearing spectacles.

4.5 Lapse detection

This section provides a summary of the key EEG-based, eye-closure-based, and video-based approaches used to date in the area of lapse detection (cf. drowsiness estimation).

4.5.1 EEG-based approaches

Sommer *et al*. (2002) applied LVQ to discriminate between clear microsleep and non-microsleep events. Four EEG channels, 2 EOG, and video were recorded from 11 young subjects while they performed a 25-min simulated driving task every hour between 1 a.m. and 7 a.m. Two cameras were used to record the driver's portrait and right eye region. Two experienced persons identified definite cases of behavioural microsleep using the video and this was used as the gold standard. Absolute spectral power in the delta, theta, alpha, sigma, and beta bands was calculated (averaged over an 8-s window, starting 4-s prior to the event). LVQ was used to discriminate between microsleep and non-microsleep, using the EEG features mentioned earlier. The authors report a best discrimination performance of 90.4 ± 1.4 %.

4.5.2 Eye-closure-based systems

The PERcentage of eyelid CLOSure over the pupil over time (PERCLOS) measure was established by Wierwille and Ellsworth (1994) and involved trained observers looking at a subject's eyes and facial changes when deciding the level of drowsiness. It was commercialized by Attention Technology Inc. (Pittsburgh, PA, USA) and uses data captured via a video camera to measure the PERCLOS value. According to the developers, PERCLOS takes slow eyelid closures ("droops") into consideration, rather than blinks. It measures the proportion of time that a subject's eyes are closed over a 1-minute period as judged by a human scorer via video recording. A subject was determined to be drowsy if their eyes were more than 80% closed. PERCLOS contains 3 drowsiness metrics: P70, the proportion of time the eyes were closed >70%; P80, the proportion of times the eyes were closed >80%; and EYEMEAS, the mean square percentage of the eyelid closure rating.

PERCLOS was determined to be the most reliable and valid measure of a person's alertness level from six drowsiness detection technologies evaluated by Dinges *et al*. (1998). They measured vigilance in subjects, who were sleep-deprived for 42-hours, via twenty discrete 20-min PVT sessions and simultaneously recorded facial video during the 42-hour period. The video was rated offline according to PERCLOS criteria by human experts. All PVT reaction times >500 ms were arbitrarily defined as performance lapses. The PERCLOS measure and PVT performance were time-locked and the coherence between the two

measures estimated. Of the six drowsiness measures, PERCLOS (P80) showed the highest correlation with 'lapses' within and between subjects (mean between-session correlation r = 0.87 ± 0.03 (mean ± SE), N = 10). It also showed a minute-to-minute correlation of 0.63 ± 0.06 between the P80 metric and PVT lapse frequency. Interestingly, this technology gave a similar mean correlation compared to an auditory vigilance task (Wierwille and Ellsworth, 1994).

4.5.3 Video-based systems

Driver State Sensor (DSS) (Seeing Machines, Canberra, Australia) is a research platform specifically developed for driver behaviour research, and consists of IR illuminators, a single video camera, and the computer vision software FaceLAB. DSS is a video-based technology which enables researchers to unobtrusively measure 3D head-pose and eye-lid motion parameters of a driver such as eye-lid closure, frequency and duration of blinks, and eye aperture in real-time. Developers claim that DSS can detect behavioural microsleep events, PERCLOS, and driver distraction using the measured video metrics. The system also has the capability to generate customized alerts when fatigue or distraction is detected. Advantages of the system include its ability to automatically calibrate the device for individuals, automatic adaptability from day to night-time driving, and its ability to cope with users who wear spectacles. There is no information available on the performance of drowsiness/lapse detection systems based on DSS/FaceLAB.

4.6 Summary

In this chapter, drowsiness was regarded as part of the arousal continuum and defined as the transitory period between alert wakefulness and sleep. The term 'lapse' was defined as a discrete event where a temporary loss of performance occurred without implying a particular state of arousal. The fact that an occurrence of a lapse does not necessarily imply that a subject's arousal level is low (and vice versa) was emphasized.

Studies investigating detection of lapses and estimation of drowsiness have utilized either auditory (Makeig and Inlow, 1993; Makeig and Jung, 1996), visual (Cajochen *et al.*, 1999; Makeig *et al.*, 2000; Van Orden *et al.*, 2000) or a combination of both auditory and visual tasks (Huang *et al.*, 2001; Jung *et al.*, 1997) to assess a subject's performance.

The gold standard for drowsiness and lapse estimation has included human experts visually rating the EEG data when there was no task (Matousek and Petersen, 1983; Subasi, 2005b; Vuckovic *et al.*, 2002), rating facial cues (Lal and Craig, 2002), performance comparison to

metrics such as local error rates (Makeig and Inlow, 1993; Makeig and Jung, 1995) and tracking error (Van Orden *et al.*, 2000).

Approaches used to estimate drowsiness include EEG analysis with measures such as power spectra (Belyavin and Wright, 1987; Jung *et al.*, 1997; Lal and Craig, 2002; Lal and Craig, 2005; Lin *et al.*, 2006; Lin *et al.*, 2005a; Lin *et al.*, 2005b; Makeig and Inlow, 1993; Makeig and Jung, 1995; Matousek and Petersen, 1983), wavelet transforms (Subasi, 2005b), combination of linear and non-linear methods (Papadelis *et al.*, 2007), and proprietary systems such as the Drowsiness Detection Algorithm by Consolidated Research Inc. Another approach has been to use eye-closure and eye-movements to estimate drowsiness in systems such as PERCLOS, Blinkometer, Alertness Monitor, Eye-Com's drowsiness detector, Optalert drowsiness measurement system, and methods proposed by Van Orden *et al*. (2000) and Grace *et al*. (2001; 1998). Head-position-based systems such as the PASS system by Advanced Safety Concepts and video-based systems (Bergasa *et al.*, 2006; Ji and Yang, 2002) are other approaches to drowsiness estimation.

In contrast to the relatively large number of drowsiness estimation approaches in the literature, there are only a few approaches that have attempted the detection of lapses. These include EEG-based approaches utilizing ANNs (Sommer *et al.*, 2002), an eye-closure-based system utilizing PERCLOS (Wierwille and Ellsworth, 1994), and a video-based system which utilizes eye-closure and head-pose information to detect lapses.

This literature review has revealed that drowsiness estimation has received greater focus from the research community than the detection of lapses, presumably due to the greater challenges posed by the latter task. Two key aspects poorly addressed to date are (1) characteristics of lapses and (2) methods for detection of lapses, particularly with high temporal resolution. Addressing these were the key research aims of this project.

4.7 Hypotheses

The research questions and corresponding hypotheses of this project were:

Hypothesis 1 – Visual identification of lapses from EEG/EOG

> **Gap in literature:** In contrast to sleep stages, it has not been established whether or not expert EEG readers are able to identify cues in the EEG and/or EOG indicating lapses.
>
> **Hypothesis:** Electroencephalographers (expert EEG readers) are able to identify cues in the EEG and/or EOG which indicate lapses.

Rationale: Human experts are able to determine drowsiness and sleep fairly accurately from the EEG and EOG (Kuwahara *et al.*, 1988) using the standard criteria (Rechtschaffen and Kales, 1968). Consequently, one would deduce that they are able to identify similar cues in the EEG and/or EOG indicative of lapses.

Significance: If hypothesis is correct, the visual cues in the EEG and/or EOG could be incorporated in the design of an automated real-time lapse detection system.

Hypothesis 2 – Lapses of responsiveness when not sleep-deprived

Gap in the literature: There is only very limited evidence that persons who are not sleep-deprived can lapse on extended tasks and no indication of the extent and characteristics of such lapsing.

Hypothesis: Normal non-sleep-deprived subjects can have multiple behavioural microsleeps of several seconds while performing extended sustained attention tasks.

Rationale: There is substantial anecdotal evidence of people lapsing during normal working hours (e.g., while driving a car or listening to a lecture). More specifically, Van Orden *et al*. (2000) found that over half of their non-sleep-deprived participants had performance lapses during a 53-min long 2-D compensatory tracking task. However, they were unable to provide any quantitative data on the incidence and duration of these lapses.

Significance: A complete loss of responsiveness (even for a few seconds) while engaged in a critical task such as driving a vehicle or landing an aircraft can have disastrous consequences in the form of serious injuries and/or multiple fatalities as well as losses to property. Information obtained about the characteristics of such lapses would be extremely valuable and will increase the current knowledge of the subject area.

Hypothesis 3 – Detection of lapses via non-linear methods

Gap in the literature: There is no evidence indicating whether or not non-linear signal processing techniques can enhance the detection of lapses over that obtainable from linear methods.

Hypothesis: Non-linear signal processing techniques can enhance the detection of lapses over that obtainable from linear methods.

Rationale: Non-linear techniques have been shown to provide information in addition to those from spectral analysis (Fell *et al.*, 1996). For example, non-linear

signal processing has been shown capable of predicting the onset of epileptic seizures several minutes in advance, despite no visible changes in the power spectra or visually in the raw EEG (Le Van Quyen *et al.*, 1999; Le Van Quyen *et al.*, 2001). Therefore, it was reasoned that the application of such techniques may be appropriate to the lapse detection problem to find 'hidden' changes in the EEG as the brain state changed from alert to non-responsive.

Significance: 'Hidden' changes in the EEG during lapses may potentially be extractable using non-linear signal processing methods. These changes may provide cues that could be incorporated into a lapse detection system. In addition, the non-linear measures may also potentially extract cues that enable lapses to be predicted.

Hypothesis 4 – Detection of lapses via EEG coherence changes

Gap in the literature: It is not known whether lapses in task performance are reflected in decreases in coherence across specific brain sites.

Hypothesis: Lapses in task performance are reflected in decreases in coherence across specific brain sites.

Rationale: It has been shown that coherence increases within and between sensory and motor sites during task performance (Aoki *et al.*, 1999; Aoki *et al.*, 2001). A relationship between increased detection error rate in an auditory vigilance task and spectral coherence in several EEG bands has also been demonstrated (Makeig *et al.*, 1996).

Significance: Coherence changes may contribute additional cues, which in turn may improve the performance of a lapse detection system.

Hypothesis 5 – Detection of lapses via changes in EEG spectral asymmetry

Gap in the literature: It is not known whether lapses in task performance are reflected as increases or decreases in spectral asymmetry between different brain sites (left versus right, and front versus back).

Hypothesis: Lapses in task performance are reflected by increased spectral asymmetry between different EEG sites (left versus right, and front versus back).

Rationale: The occurrence of EEG desynchronization between different brain sites has been reported preceding seizures (Mormann *et al.*, 2003). Similarly, EEG activity

in different cortical regions may desynchronize during lapses. This in turn may be reflected as increased spectral asymmetry between different EEG sites.

Significance: Changes in spectral asymmetry may contribute additional cues which may enhance a lapse detection system.

CHAPTER 5

Rating Electrophysiological Data for Vigilance Lapses

5.1 Introduction

As mentioned in Chapter 1, it would be of immense value if lapses could be detected automatically and in real-time, so that preventative action could be taken to avoid accidents.

Automating the detection process by observing cues in the EEG and EOG indicative of lapses is one approach that could be undertaken. This process can be simplified considerably by designing a system that could mimic human experts; that is, similar to the approach often taken in the automation of the detection of epileptiform activity in the EEG [e.g., (Dingle *et al.*, 1993)]. Surprisingly, the extent to which human EEG raters can detect lapses is unknown. Therefore, as a first step, it was decided to determine the ability of human expert EEG readers to detect lapses. Consequently, a study was undertaken in which 4 expert EEG raters classified the alertness level of a 10-min record from each of 10 air traffic controllers (ATCs), using cues in the EEG and EOG.

5.2 Methods

The ATC data used in this rating study was collected by Dr. Leigh Signal as part of her research study into the efficacy of scheduled napping during the night shift on the performance and neurophysiological alertness of ATCs in an operational environment (Signal, 2002; Signal and Gander, 2007).

The fundamental task of an ATC is to ensure the safe and efficient passage of aircraft both on the ground and in the air. This involves the controller working in a monitoring role within an automated environment. Airways Corporation of New Zealand Ltd is solely responsible for providing air traffic control services in New Zealand. Signal's study involved ATCs working in the Christchurch Radar Centre as Area Controllers. They were responsible for the upper-level airspace (above 3000 m) across the majority of New Zealand and down to the Antarctic.

5.2.1 Subjects

Thirty five ATCs were approached and 28 (19M/9F; mean age 35.5 years, range = 26–56) agreed to participate in the study (response rate = 80%).

5.2.2 Neurophysiological measures

Subjects were connected to an ambulatory neurophysiological recorder (*Embla*™ by Flaga hf, Reykjavik, Iceland) and five EEG channels recorded: C4-A1, O2-A1, O2-Oz, O2-P4, and Oz-P4. This configuration allowed the collection of data from midline and temporal locations as well as central and occipital sites.

Left and right EOG were recorded from the left and right outer canthus, with the left electrode positioned 1 cm up from the horizontal plane, and the right 1 cm down, with both referenced to the auricular reference (A1). This placement allowed the recording of both horizontal and vertical eye movements and the identification of electrode artifact in the EEG.

A number of factors affected channel selection. These included the ability to detect slower EEG frequencies associated with decreased alertness, allowing the scoring of sleep, minimizing known causes of artifact, and including both widely spaced referential and closely placed bipolar channels. Ideally, a full 10-20 montage would have been desirable but was not feasible given the time constraints associated with connecting such a montage Therefore 5 EEG channels were used as a compromise (Signal, 2002).

Channel C4-A1 was used to score sleep according to standard criteria (Rechtschaffen and Kales, 1968). The choice of location of the three bipolar channels was based on the need to detect changes in alpha and theta with increasing drowsiness. During drowsiness, the alpha rhythm is prominent in the occipital area and theta activity is generally seen posteriorly (Niedermeyer, 1999). These rear electrodes are also less likely to record eye movement artifacts and their use provided a combination of referential and bipolar channels within the montage.

Bipolar electromyogram (EMG) recorded from two electrodes positioned on the mentalis/submentalis muscles (muscles beneath the chin) was used to identify muscle artifacts in the EEG. Similarly, electrocardiogram (ECG) was recorded to identify cardiac artifacts.

The analogue channels were sampled at 2000 Hz and digitized to 16-bit resolution using a Sigma-Delta A-D converter. It was then down-sampled to 200 Hz, and digitally pass-band filtered at 0.5–90 Hz before storage.

5.2.3 Visuomotor performance measure

The psychomotor vigilance task (PVT) (Dinges and Powell, 1985) was used to assess the performance of the ATCs across the night shift. It is a validated, reliable, and sensitive test of vigilance and simple visual reaction time (Dinges *et al.*, 1998). It was chosen for the ATC study due to its "known psychometric properties, ease of use in an operational context, and lack of practice effects" (Signal, 2002).

The test required individuals to respond as fast as possible to the presentation of digits on an LED display. The test was 10 min in duration and the inter-stimulus interval range was 2–10 s. This resulted in approximately 100 reaction stimuli per test. The voltage difference between the stimulus onset and the corresponding reaction by the subject resulted in a square-wave signal, which was fed into the *Embla*™ recorder, and recorded simultaneously with the physiological data channels. The reaction time (RT) is the duration from stimulus onset to when the subject responds to the stimulus by pressing the button.

Subjects were free to choose to use either their index finger or thumb to push the response key but were required to use the same method of responding for all tests.

5.2.4 Experimental procedure

Each ATC was involved in the study on four separate occasions. Two of these involved night shifts starting at 22:30 and two at 23:30. For a given shift start time, the controller was asked to take a 40 minute nap during one session, and to remain awake during the other. This allowed the effectiveness of a nap to be assessed on both types of night shifts as part of Signal's research.

All subjects completed the PVT three times (mid-way through the shift, during a scheduled break, and at the end of the night shift) during each study night shift, giving a total of 12 PVT sessions per subject. Each subject completed a 1-min trial of the PVT prior to the study. Each test was conducted in a room without distractions and the subjects were asked to hit a

response key as quickly as possible on presentation of numbers in an LED display. The resulting display indicated the response time in milliseconds, giving the subject feedback on their performance. The subject's EEG and EOG were recorded while they performed the PVT.

5.3 Rating study

All RTs greater than 500 ms were arbitrarily defined as *lapses* (Dinges *et al.*, 1997). Note that this particular definition of 'lapse' only applies to this chapter of the thesis.

A pilot study involving one expert rater, scoring the alertness level of 5 ATCs based on 10 min of EEG while they performed a PVT, showed that the expert was unable to detect lapses by observing the EEG. It was decided to undertake a more comprehensive study involving 4 raters analysing 10 independent subjects who lapsed at least once during a 10-min session from the ATC database.

5.3.1 Objectives

The objectives of the study were to (1) determine whether experts could reliably identify lapses in visuomotor performance by observing cues in the EEG and/or EOG, and (2) determine the relationship between reaction time and level of arousal.

5.3.2 Data

The ATC database containing physiological and performance data recorded from 28 ATCs was utilized for the rating study. The database consisted of 304 sessions collected from 28 ATCs with a total of 430 lapses. Thirty two sessions were excluded due to various reasons such as poor data quality. A substantial number of sessions (184) did not contain any lapses. All sessions were ranked according to the number of lapses, and the 10 sessions from independent subjects containing the highest number of lapses were selected to be rated by human experts. The 10 sessions selected for the rating study contained 101 of the 430 lapses.

Three physiological data combinations (EEG only, EOG only, EEG+EOG) were derived for each chosen session of all 10 subjects to produce a total of 30 records (total of 5 h of data).

It was verified that the selected subjects were alert during the first 10 s of the PVT by inspecting the RTs of each session to ensure that they did not contain any lapses during the first 10 s. This provided the raters with a baseline alertness level for each subject.

5.3.3 Rating procedure

Two teams of 2 raters each were asked to classify the records using a continuous marker into one of five categories: 1) Alert, 2) Light Drowsy, 3) Deep Drowsy, 4) Lapse in responsiveness (i.e., complete, temporary loss in responsiveness) other than sleep, and 5) Sleep. The first rating team had particular expertise in clinical neurophysiology and the second team in sleep research.

Before commencing the rating, each team was required to define and write down a single set of criteria (agreed by both raters) they would use to identify the various levels of the rating scale. If the rating team felt that there was no basis or that there are no cues on which to base a 'lapse in responsiveness other than sleep', they were instructed to ignore the level and make a written statement that they could not estimate the presence of such lapses from the EEG and/or EOG. Other than the consultation described, each expert rated the data independently of the three other raters.

The records were presented to the raters in a random order by a program developed by the author using *MATLAB*™. A screenshot of the rating program is shown in Figure 5-1. Each rater received the records in the same sequence. Raters were blind to the PVT performance of the subject.

Data from two additional subjects were provided to the raters for the purpose of training, prior to the rating of the 10 test subjects. This data was used by the rating teams to set up and refine team rating rules as well as to familiarize themselves with the rating software. Once training was complete, each rater proceeded to rate the 10 test subjects according to the pre-defined criteria.

5.4 Analysis

5.4.1 Detecting lapses

The first objective of the rating study was to determine whether expert raters could identify lapses using the EEG and EOG. A rater scored a *hit* if the rating was marked as '4' (lapse) or '5' (sleep) for some period during a PVT lapse. Conversely, it was classified as a *false positive* (FP) if there was no lapse during the period the rater classified as '5', provided it did not occur during an inter-stimulus period. If this was the case, one could not definitively conclude that the rater made a false detection as there was no performance sampled during the inter-stimulus period. The number of hits and FPs were calculated for EEG only, EOG only, and EEG+EOG rating data.

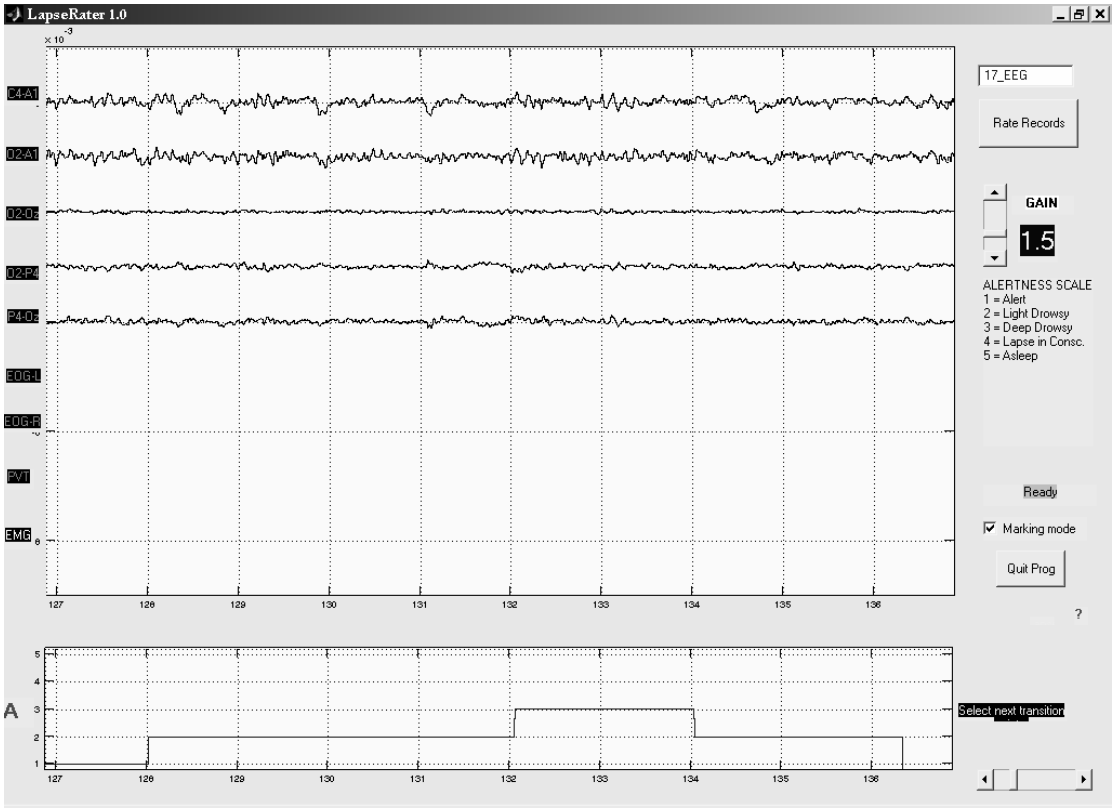

Figure 5-1 A screenshot of the software used for the rating study. The top section displays the EEG, EOG, or both. The expert was asked to mark transitions on the bottom axis. Please note that the rating shown in the figure is simulated for illustration purposes only.

5.4.2 Level of arousal and reaction time

The second objective of the rating study was to determine the relationship between arousal level (as rated by experts observing the EEG and EOG) and reaction time.

For a given subject, the correlation coefficient (r) between an expert's rating (at stimulus onset) and RT was calculated. Mean r across 4 raters was calculated for each subject. The t-test was used to determine if there was a significant correlation between the rated level of arousal and RT.

Similarly, the correlation between variability of reaction times and level of arousal rating was calculated. Variability was determined by calculating the standard deviation of RTs in each 1-min epoch. The rating value for an epoch was the mean rating across the epoch. The standard deviation of RTs in the epoch and the mean rating for a given rater were used to calculate a correlation coefficient for a subject. The mean correlation for a subject was obtained by averaging across the 4 raters. The t-test was used to determine if a significant correlation existed between the two variables.

This procedure was repeated for the 3 rating types.

5.4.3 Inter-rater agreement

Initially, percentage agreement between rater pairs was used as a measure of inter-rater agreement. However, this is a less than ideal measure due to it being obscured by chance agreement between the raters.

The *kappa statistic* is considered to be an improvement over percent agreement as it factors out chance agreement (Everitt, 1996). This statistic is one of the most commonly used measures to determine inter-rater agreement when observing qualitative/categorical variables and is often referred to as 'Cohen's kappa' (Cohen, 1960).

A succinct explanation of Cohen's kappa calculation is provided by Mielke and Berry (2001). It is assumed that two raters independently classify each of g observations into r discrete, mutually exclusive, and exhaustive classes. The resulting classifications can be summarized on an r by r cross-classification box as shown in Table 5-1, with proportions as cell entries.

Cohen's kappa is given by

$$\kappa = \frac{P_O - P_E}{1 - P_E} \tag{5-1}$$

where $P_O = \sum_{i=1}^{r} P_{ii}$, $P_E = \sum_{i=1}^{r} (P_{1i} + P_{2i} + ... + P_{ri})(P_{i1} + P_{i2} + ... + P_{ir})$, and elements in the cross-classification box sum to 1. Note that P_O is the observed proportion of observations for which the two observers agree, while P_E is the proportion of observations which is expected to agree by chance alone. Therefore, $P_O - P_E$ is the proportion of agreement beyond what is expected by chance and $1 - P_E$ is the maximum possible proportion of agreement beyond what is expected by chance. Thus, the coefficient κ is the proportion of agreement between the two raters after the removal of chance agreement (Mielke and Berry, 2001).

Kappa coefficient values are constrained to lie in the interval [0,1]. If there is complete agreement between the raters, then $\kappa = 1$. If the observed agreement is greater than chance, then $\kappa > 0$. In the "unlikely event of the observed agreement being less than chance, $\kappa < 0$, with its minimum value depending on the marginal distributions of the two raters" (Everitt, 1996). Landis and Koch (1977) provided an arbitrary benchmark for the interpretation of the observed κ values, as shown in Table 5–2. The Kappa statistic has been used to assess the inter-rater reliability in sleep-stage scoring, including a collaborative study between 8 European sleep laboratories participating in the SIESTA project, with a large number of patients with different sleep disorders (Danker-Hopfe *et al.*, 2004).

Table 5-1 An example of a 3 by 3 cross-classification (Mielke and Berry, 2001). Note that the cell entry P_{12} refers to the proportion of observations which the first rater classifies has belonging to class 1, and the 2nd rater classifies as belonging to class 2, etc.

	Columns			
Rows	**1**	**2**	**3**	**Sum**
1	P_{11}	P_{12}	P_{13}	P_{1*}
2	P_{21}	P_{22}	P_{23}	P_{2*}
3	P_{31}	P_{32}	P_{33}	P_{3*}
Sum	P_{*1}	P_{*2}	P_{*3}	P_{**}

Table 5-2 An arbitrary benchmark for interpreting kappa values by Landis and Koch (1977).

Kappa (κ)	Strength of agreement
0.00	Poor
0.01–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Perfect

One rater, r1, was selected as the reference for the alertness level. The three remaining raters (r2, r3, and r4) were taken and the kappa statistic calculated between all combinations of rater pairs using r1 as the reference for each level of alertness. This process was repeated until all raters were used as the reference. This procedure was repeated across all subjects. For a given alertness level, the mean kappa across the 4 raters was calculated for each subject. This yielded 10 mean kappa values per alertness level and these values were plotted on a box plot.

5.5 Results

5.5.1 Reaction time profile

Reaction times were divided into 'non-lapses' (RT <500 ms) and 'lapses'. The duration (mean ± SE) of non-lapses was 0.308 ± 0.013 s and the mean duration of lapses was 1.002 ± 0.121 s across all subjects. There were a total of 101 lapses in the 10 subjects studied. The mean number of lapses across all subjects was 10.1 (range 2–31). The longest lapse was 11 s. The majority of lapses (62%) in the dataset were less than 1 s.

Figure 5-2 shows the mean and standard error of the duration distributions of all RTs.

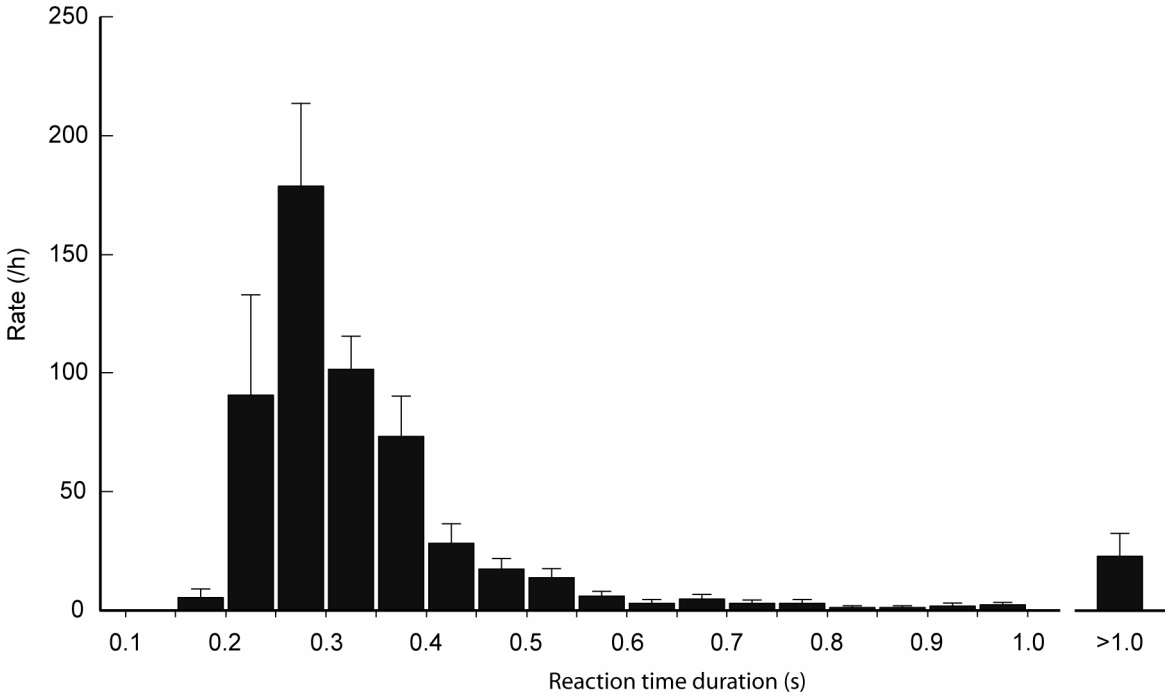

Figure 5-2 Histogram depicting the distribution of the duration of RTs across all subjects. Standard error bars are shown.

The mean RT and the variance of RTs in each 1-min epoch were calculated. Figure 5-3 shows the time course of the mean duration and variance of RTs across the 10-min PVT test and the standard error. Both RT duration and variance show an increasing trend over the 10-min period.

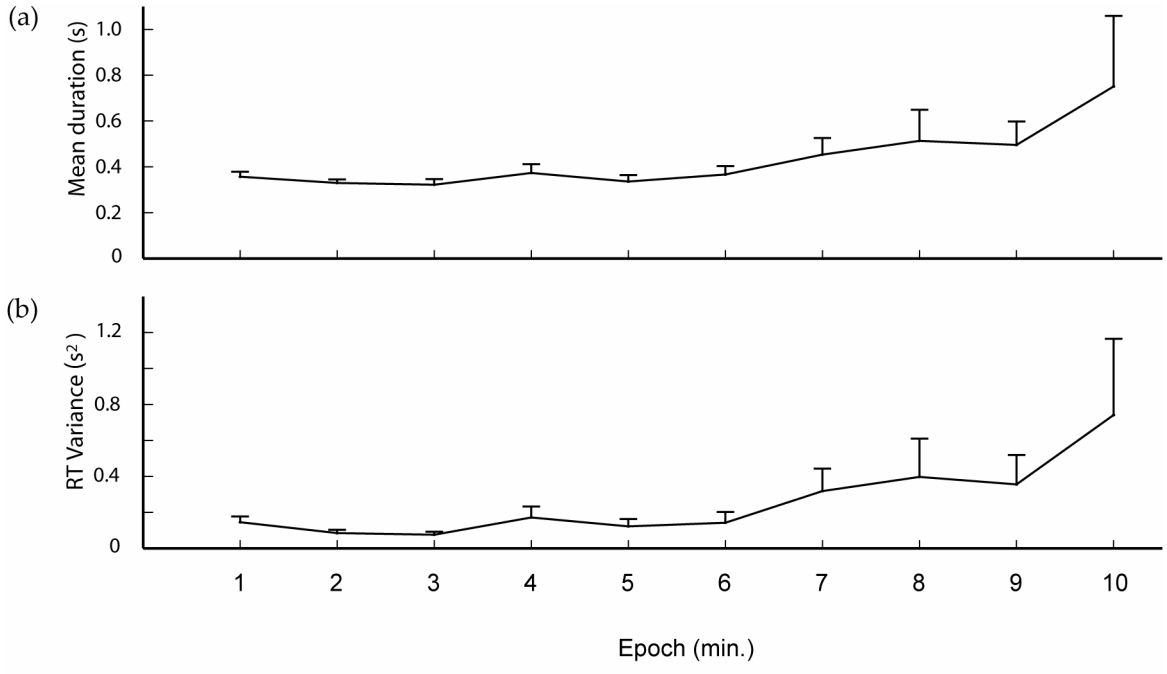

Figure 5-3 (a) Mean RT duration in each 1-min bin across all subjects (*N*=10). (b) Variance in RTs across the bins. Standard error bars are shown for both cases.

5.5.2 Detection of lapses

None of the raters identified a category '4' (i.e., lapse in responsiveness) event as both teams agreed prior to the rating that there was no basis on which to base a lapse of responsiveness. Therefore, this category was ignored for the remainder of the analysis.

Table 5-3 shows the number of lapses as determined by the PVT and the numbers of hits and false detections made by each rater using cues in both the EEG and EOG. A total of 6 lapses were correctly detected by at least one rater (duration 0.685, 1.315, 1.905, 4.470, 5.895, 11.460 s) but with agreement between two raters on just 2 of these lapses (duration 1.315, 5.895 s). No lapses were detected by raters in 8 of the 10 subjects.

For the case of EEG only, 7 lapses were correctly detected by one or more raters (durations 0.535, 1.315, 1.595, 1.610, 1.905, 4.470, 5.895 s). There was agreement between 2 different rater pairs on 2 lapses (durations 1.905, 5.895 s). These results are summarized in Table 5–4. No lapses were detected in 9 of the 10 subjects.

There were 4 lapses in common between the lapses detected using EEG+EOG and lapses detected using EEG-only (durations 1.315, 1.905, 4.470, 5.895 s). In the EEG+EOG case, the period immediately preceding 4 of the 6 lapses was classified as EEG-microsleeps. One of the two remaining lapses was classified as 'light drowsy', and the other was classified as 'deep drowsy' by one rater and 'sleep' by another. In the EEG-only case, the period immediately preceding 6 of the 7 lapses was classified as 'sleep' and the remaining lapse was classified as 'alert'.

Table 5-3 Lapse detection breakdown for the 4 raters marking alertness level of subjects simultaneously viewing EEG and EOG.

Subject	No. of	Rater 1		Rater 2		Rater 3		Rater 4	
	lapses	Hits	FP	Hits	FP	Hits	FP	Hits	FP
1	31	0	0	0	0	2	1	0	0
2	21	0	0	0	0	2	0	4	0
3	17	0	0	0	0	0	0	0	0
4	12	0	0	0	0	0	0	0	0
5	5	0	0	0	0	0	0	0	0
6	4	0	2	0	0	0	0	0	0
7	4	0	0	0	0	0	0	0	0
8	3	0	0	0	0	0	0	0	0
9	2	0	0	0	0	0	0	0	0
10	2	0	0	0	0	0	0	0	0
Total	**101**	**0**	**2**	**0**	**0**	**4**	**1**	**4**	**0**

Table 5-4 Lapse detection breakdown for the 4 raters marking alertness level of subjects viewing EEG only.

Subject	No. of lapses	Rater 1		Rater 2		Rater 3		Rater 4	
		Hits	FP	Hits	FP	Hits	FP	Hits	FP
1	31	0	0	0	0	0	0	0	0
2	21	0	0	0	0	3	0	6	3
3	17	0	0	0	0	0	0	0	0
4	12	0	0	0	0	0	0	0	0
5	5	0	0	0	0	0	0	0	0
6	4	0	2	0	0	0	0	0	0
7	4	0	0	0	0	0	0	0	0
8	3	0	0	0	0	0	0	0	0
9	2	0	0	0	0	0	0	0	0
10	2	0	0	0	0	0	0	0	0
Total	**101**	**0**	**2**	**0**	**0**	**3**	**0**	**6**	**3**

None of the raters had hits or FPs using the EOG alone. Therefore no further analysis was performed on EOG-only ratings.

5.5.3 Level of arousal and reaction time

Figure 5-4 (a) shows the mean correlation coefficients between EEG+EOG rating level and RT duration, and Figure 5-4 (b) shows the mean correlation between RT variance and rating. The *r*-values across subjects were not significantly different from zero for RT duration (*t*-test, $p = 0.3102$) or RT variance (*t*-test, $p = 0.7054$). This indicates that there was no overall correlation between the level of arousal (as indicated by the expert rating of the EEG+EOG) and reaction time.

Similarly, no significant correlations were found between RT duration/variance and EEG only or EOG-only ratings.

5.5.4 Level of arousal and rate of lapsing

The overall proportion of lapses beginning at each arousal level (rated by observing the EEG and EOG) was calculated. The assigned level at the start of most lapses was Alert (mean proportion across raters = 76%) and far fewer started during epochs marked as Light Drowsy (14%), Deep Drowsy (8.4%), or Sleep (1.5%). Since Alert was the most common rating level, average lapse rates were also calculated within each rating level. The total number of lapses within a level was divided by the total duration of that level. This analysis was applied across data from all sessions and was completed separately for each rater. The mean lapse rates across raters increased with decreasing level of arousal (as ascertained from EEG). Mean ± SD lapse rate was 1.01 ± 1.04 lapses/min for Alert, 1.16 ± 2.00 lapses/min for Light

Drowsy, 1.52 ± 2.84 lapses/min for Deep Drowsy, and 6.76 ± 6.11 lapses/min for Sleep (Figure 5-5).

Out of the 101 lapses observed over the 10 subjects, 33 were rated as 'alert' by all 4 raters (~33%). Figure 5-6 shows the distribution of the RTs of 'alert' lapses. Of the 33 lapses rated as 'alert' by all 4 raters, 21 lapses were less than 1.0 s duration, 8 lapses were 1–2 s, and 4 lapses were >2.0 s.

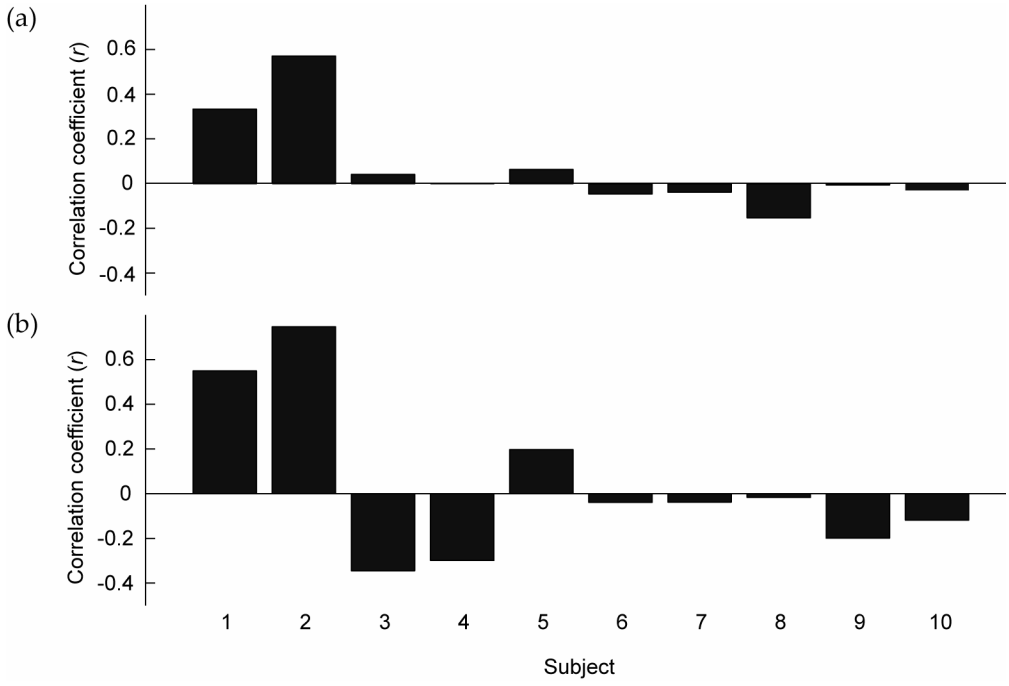

Figure 5-4 The mean correlation values between (a) RT duration and EEG+EOG rating value at the start of the RT, and (b) Variance of RTs in 1-min epochs and the mean EEG+EOG rating value for the epoch.

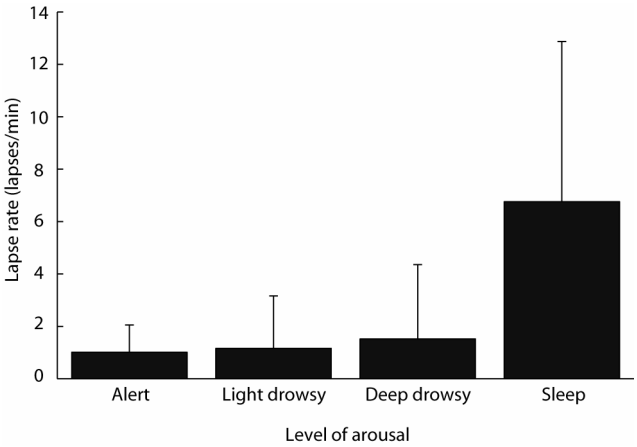

Figure 5-5 Level of arousal (ascertained by 4 raters observing the EEG) *vs*. mean lapse rate. The SD is also shown.

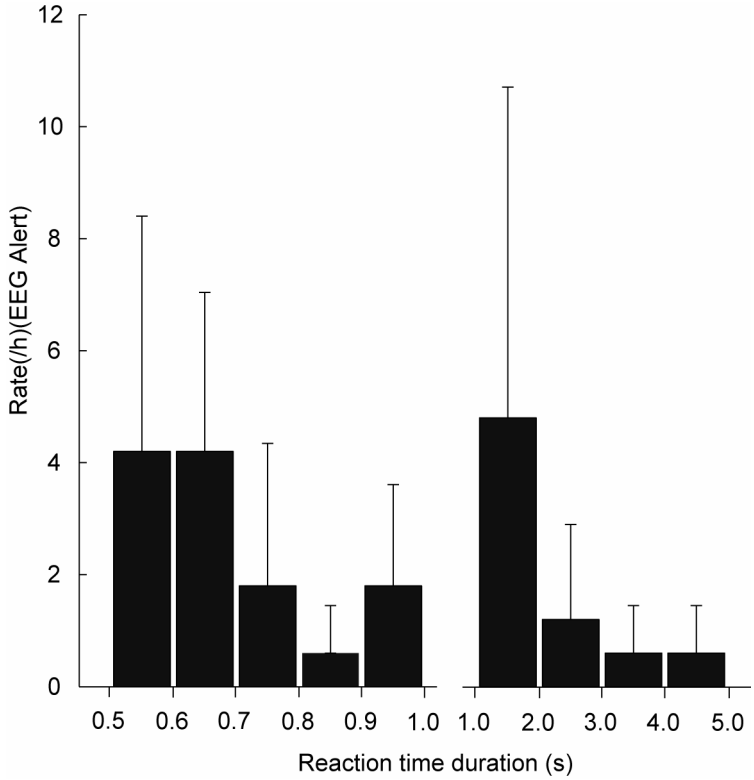

Figure 5-6 Histogram depicting the distribution of lapses that occurred during periods for which all 4 raters classified the subjects' state as 'alert'. Standard error bars are shown.

5.5.5 Inter-rater agreement

The inter-rater reliability of the human ratings was calculated by way of mean sample-by-sample percent agreement between all rater pair combinations and the results are shown in Table 5-5. Percent agreement varied widely from 27–89% for the EEG-only rating, 68–92% for EOG-only, and 53–82% for EEG+EOG. The percentage of the total data categorized into each arousal level by each of the four raters, reading EEG-only, EOG-only, and EEG+EOG data, are shown in Tables 5–6 to 5–8. There was unanimous rating agreement only for the 'Alert' stage (25% of total time when rating EEG-only, 56% when rating EOG-only, and 42% when rating EEG+EOG).

Table 5-5 Mean percent agreement between rater pairs for the EEG only, EOG only and EEG+EOG data (across all subjects).

	Rater Pairs					
	GC/PP	**GC/LS**	**GC/MV**	**PP/LS**	**PP/MV**	**LS/MV**
EEG	44%	32%	27%	74%	78%	89%
EOG	73%	68%	68%	92%	76%	75%
EEG+EOG	58%	53%	60%	75%	82%	80%

However, percent agreement is not the most appropriate measure as it is obscured by chance agreement between the raters. Therefore, the kappa statistic was used as a measure of inter-rater agreement for the EEG+EOG rating. Figure 5-7 shows the box plots of kappa values for various levels of alertness for the (a) EEG+EOG rating and (b) EEG-only rating case. According to Landis and Koch's (1977) interpretation (see Table 5–2), these kappa values correspond to a 'poor' to 'slight' level of agreement between raters, at all alertness levels. The mean inter-rater agreement in the study was poor for all rating levels. For example, the kappa value for the 'alert' stage was around 0.04.

Table 5-6 Percentage of data categorized into each arousal level by each rater (reading EEG-only). The mean across the raters is also shown.

	Level of Arousal			
Rater	**Alert**	**LD**	**DD**	**Sleep**
1	29.0%	69.9%	1.0%	0.0%
2	84.2%	15.1%	0.7%	0.0%
3	89.2%	7.9%	2.7%	0.2%
4	93.5%	3.5%	2.4%	0.6%
Mean	**74.0%**	**24.1%**	**1.7%**	**0.2%**

Table 5-7 Percentage of data categorized into each arousal level by each rater (reading EOG-only). The mean across the raters is also shown.

	Level of Arousal			
Rater	**Alert**	**LD**	**DD**	**Sleep**
1	73.2%	26.8%	0.0%	0.0%
2	100.0%	0.0%	0.0%	0.0%
3	91.6%	6.0%	2.4%	0.0%
4	76.2%	17.9%	5.9%	0.0%
Mean	**85.2%**	**12.7%**	**2.1%**	**0.0%**

Table 5-8 Percentage of data categorized into each arousal level by each rater (reading EEG+EOG). The mean across the raters is also shown.

	Level of Arousal			
Rater	**Alert**	**LD**	**DD**	**Sleep**
1	65.5%	33.9%	0.5%	0.1%
2	91.0%	9.0%	0.0%	0.0%
3	78.4%	15.8%	5.6%	0.2%
4	90.6%	5.4%	3.6%	0.4%
Mean	**81.4%**	**16.0%**	**2.4%**	**0.2%**

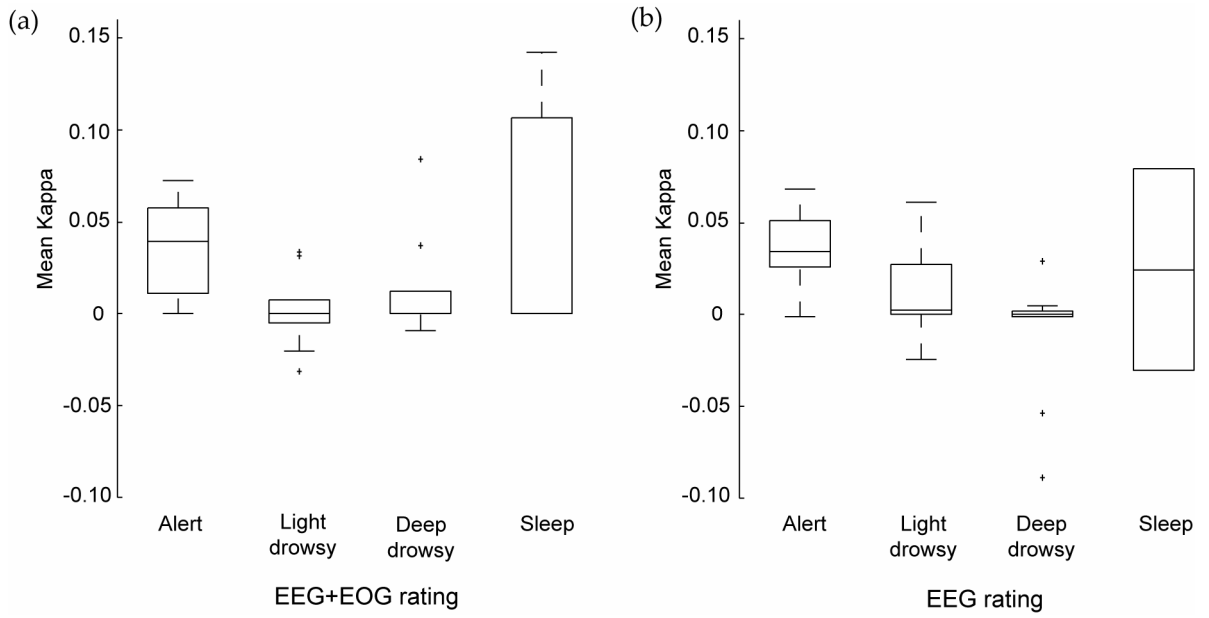

Figure 5-7 A modified 'box and whisker' plot of mean kappa values for each alertness level for (a) EEG+EOG rating and (b) EEG rating. The central box spans the first and third quartiles; the line in the box marks the median value, and lines extend from the box out to the smallest and largest values that are not suspected outliers (defined as observations more than 1.5 times the inter-quartile range). These points are plotted individually on the box and whisker plot (Moore and McCabe, 2003).

5.6 Discussion

Detection of vigilance lapses from EEG/EOG by human expert raters was very low, with only 6 of 101 lapses detected while viewing the EEG and EOG simultaneously and only two of these concomitantly by two raters. Slightly more (7) lapses were detected viewing EEG alone and none were detected when viewing the EOG alone, suggesting EOG was of little value. Lapses occurred more frequently during intervals rated as Sleep or Deep Drowsy but, since most of the data was rated Alert, the majority of lapses occurred when subjects were physiologically alert.

Even though only those sessions containing the most lapses were selected, more than half the selected subjects had less than five lapses during the entire session. The majority (62%) of the 101 lapses were less than 1 s. These results suggest that there are no overtly visible cues in the EEG and EOG during these short lapse events. Raters are thus unable to detect them consistently and, in most cases, not at all.

There were several limitations in the ATC dataset which may have hindered the rating study outcomes:

1. The ATC dataset contained RTs measured at discrete points (varying randomly between 2–10 s) as its performance measure. An accurate estimate of lapse duration could not be obtained because PVT stimuli appeared at random discrete intervals and, as a result, performance could not be assessed during the inter-stimulus period. Consequently, it is possible that the actual lapse commenced several seconds in advance of the stimulus onset but was only apparent after the occurrence of the stimulus and the consequent prolonged RT. For this reason, ratings of 'sleep' occurring during the inter-stimulus interval were not counted as FPs.

2. Psychomotor tests often make heavy demands on the subject, thereby raising the level of cerebral activity, which can temporarily mask any possible signs of fatigue [(Grandjean, 1979) as cited in (Lal and Craig, 2001)].

3. The short duration (10 min) of the PTV test may account for the limited number of lapses in the dataset. Time-on-task increases the probability of lapses (Doran *et al.*, 2001). The test length may have been too short for a subject to exhibit long lapses, as they could have forced themselves to be alert for a short period of time, despite being tired. A longer test duration may have lead to larger variations in the arousal state of subjects.

4. The definition of a lapse as RT >0.5 s was arbitrary. The majority of lapses (62%) were less than 1.0 s. Many or all of these shorter lapses may have been prolonged reaction times due, say, to drowsiness rather than complete lapses of responsiveness (e.g., behavioural microsleep) – although the exact distinction between a prolonged RT and a lapse is debatable.

5. Another important limitation of the dataset was the lack of an independent measure of alertness, such as video of the face and/or eyes. This made it impossible to determine unequivocally whether any of the lapses were due to behavioural microsleep, a lapse of sustained attention, or a lapse due to diverted attention.

6. Although sufficient for sleep rating, the EEG data only contained 5 asymmetrically distributed posterior EEG channels. Clinical neurophysiology raters (cf. sleep researchers) are tuned to interpretation of full-head EEG and would likely have found the task more difficult with fewer EEG channels and ones which provided limited sampling over the head. For example, there were no frontal EEG channels; hence no eye movement cues during the EEG-only rating.

7. Similarly, there were ambiguities in distinguishing vertical and horizontal eye movements resulting from the location of the EOG electrodes. A mixture of both

horizontal and vertical movements in the EOG channels may have prevented the raters from unequivocally distinguishing different eye movements.

The mean inter-rater agreement in the study was poor for all rating levels. For example, the kappa value for the 'alert' stage was around 0.04, markedly lower than the 0.8 which Danker-Hopfe *et al.* (2004) observed for the same stage. There are several possible reasons for this lower-than-expected mean inter-rater agreement. The rating teams were allowed to define their own criteria for rating the ATC data. It was foreseen that this allowance would invariably lead to less inter-rater agreement. However, since the rating teams had different EEG scoring backgrounds, it was to be determined if one group had a superior set of criteria for the detection of vigilance lapses. This approach was in contrast to typical sleep rating studies which use the 'gold', albeit limited, standard as defined by Rechtschaffen & Kales (1968). Neither team was found to have a substantially superior detection rate than the other. The second rating team (sleep research background) detected 7 of the 101 lapses between them when rating the EEG only and 6 when rating EEG+EOG, whereas the first team (clinical neurophysiology background) did not detect any in both rating cases.

A substantial proportion (33 out of 101) of lapses occurred when all 4 raters marked the level of arousal of the subject as 'alert' (by observing the EEG), indicating that they occurred reasonably frequently. However, some 'alert lapses' could be attributed to factors unrelated to the level of arousal. Researchers have identified factors such as distraction, a decrease in the amplitude of the finger movement with respect to the button, or the loss of a stimulus during an eye blink, to cause lapses during the PVT (Priest *et al.*, 2001). Figure 5-6 shows that the durations of these 'alert lapses' were not limited to a relatively low range (say 0.5–0.6 s). This again emphasizes the fact that the PVT does not allow one to easily differentiate between lapses of responsiveness and sluggish RTs due to the subject being, say, deep drowsy. Therefore, one cannot definitively conclude that subjects had a substantial number of *unequivocal* lapses, while appearing to be alert, from this data. Conversely, with eight RTs >1.0 s and four >2.0 s, there is strong evidence of a small number of substantial lapses when subjects were alert EEG-wise. Whether these are behavioural microsleeps or lapses of sustained attention cannot be determined from the current data. As far as the author is aware, this is the first study to demonstrate the presence of extended lapses (>1.0) in EEG-alert normal subjects.

Despite these shortcomings, these results demonstrate that detecting lapses based on EEG and/or EOG is, for the most part, *not* possible even by expert EEG raters. This suggests that, in contrast to an automated sleep stager, an automated lapse detection system will most likely need to identify features not visible in the raw EEG.

CHAPTER 6

Continuous Tracking Task Study: Design and Methods

6.1 Introduction

6.1.1 Overview

As mentioned in § 5.6, there were several limitations in the ATC data. Therefore, a new study was conducted to collect physiological and performance data from human subjects. This allowed customization of the experimental methodology to suit the requirements of the lapse project. It was hoped that an appropriately designed experimental paradigm would induce lapses in subjects and allow capture of these events. Data collected would provide invaluable spatial, temporal, and state information on physiological processes during, before, and after brief lapses. A number of signal processing techniques were applied to the physiological data from this study with the aim of finding reliable detectors and predictors of lapses.

6.1.2 Experimental requirements

There were five key requirements in the study:

1. An experimental paradigm to maximize the likelihood of subjects lapsing during normal working hours, without resorting to sleep deprivation, depressants, or excessive consumption of food.
2. Continuous sampling of a subject's performance.
3. Recording of full-head EEG and separate horizontal and vertical EOG.

4. A continuously changing target/stimulus to allow changes in performance to be detected and quantified with high temporal accuracy.
5. Incorporation of an independent measure of lapses (such as video) to allow the verification of the observations in the primary performance indicator.

6.1.3 Study hypothesis

It was hypothesized that (a) at least some subjects undertaking a continuous tracking task for a reasonably long period of time and in an environment conducive to sleeping would have occasional lapses, (b) there would be physiological cues during and prior to lapses in the EEG or eye movements, (c) these cues could be detected with reasonable accuracy using one or more signal processing techniques, and (d) the detection of physiological cues could be used in the transport industry to help avoid serious accidents.

6.2 Methods

6.2.1 Subjects

Fifteen normal healthy male volunteers aged 18–36 years (mean = 26.5) were recruited for the study. The purpose of the age range and gender restriction was to limit the sources of variation in the data. Subjects were recruited via e-mail advertisements at the University of Canterbury and Department of Medical Physics and Bioengineering, Christchurch Hospital. None had a current or previous neurological or sleep disorder and all had visual acuities of 6/9 (= 20/30) or better in each eye. All subjects considered that they slept normally the previous night (mean = 7.8 h, SD = 1.2 h, min = 5.1 h) and, hence were considered non-sleep-deprived. They were required to have at least a basic comprehension of the English language. All subjects provided informed consent prior to participating in the study. Ethical approval for the study was obtained from the Canterbury Ethics Committee.

6.2.2 Apparatus

The long-term EEG (LTEEG) monitoring system in Christchurch Hospital's Neurology Department was borrowed for the study. This system had been designed to monitor patients with suspected epilepsy and to capture significant EEG events over several days. Amplifier gain settings were modified to enable the recording of EOG signals which had higher amplitudes than the EEG. Table 6-1 summarizes the specifications of the *SMCTests*™ hardware and software used in the CTT study and Table 6-2 summarizes the hardware and software specifications of the LTEEG system.

Table 6-1 Hardware and software specification of *SMCTests*™.

PC system
Intel Pentium III 733 MHz
128 MB RAM
2 VGA monitors (Philips 17″ 107S Lightframe™ running the *SMCTests* program and the Compaq 17″ S710 displaying the target waveform). The Compaq monitor was set to 800 X 600 resolution.
Primary and secondary graphics cards: AGP/PCI
Advantech PCL 1710 data acquisition board (PCI)
Sampling rate – 64 Hz
Specialized hardware (test input devices)
Steering wheel (395 cm diameter)
SMCTests analog interface box
Software
Windows 98 (SE)
Microsoft Data Access Objects library 3.0/3.5
DirectX 8.1 runtime
SMCTests v4.3

Table 6-2 Hardware and software specification of the long-term EEG monitoring system (Dove, 2000).

Inputs
20 channels (19 referential inputs, 1 bipolar input)
Grounding – linked ears
Reference point – forehead
Isolation – BF (provided by patient room interface unit)
Amplification – instrumentation amplifier input followed by two stages.
Total analog gain – 1280
High CMRR (> 100 dB)
Bandwidth 0.1–100 Hz (20 dB/decade high and low-pass)
A-D converters
Configuration – 5 converters, each performing 4 conversions/sample
Type – 16 bit serial, bipolar
Sensitivity – 1 LSB = 80 nV, Full scale = ±2.6 mV
Sampling rate – 256 Hz
Impedance checking
Configuration – current injection into each electrode relative to forehead
Current injected – 90 nA p-p square wave symmetrical about zero.
Software
Language – Microsoft C++ version 5.0
Platform – PC, Windows 95

6.2.3 Physiological measures

Full-head EEG was recorded from 16 referential channels (Fp2, F4, C4, P4, O2, Fp1, F3, C3, P3, O1, F8, T4, T6, F7, T3, and T5) placed according to the 10/20 international system (see Figure 2-3). It was considered important to sample EEG from most of the cerebral cortex (Wright *et al.*, 1995). For example, Cajochen *et al*. (1999) found that the frontal areas of the brain are more susceptible to sleep loss than the occipital areas and considered "frontal EEG activity and ocular parameters may be used to monitor and predict changes in neurobehavioural performance associated with sleep loss and circadian misalignment." (Frontal EEG channels were not present in the ATC data). Additionally, the symmetric layout of the electrodes made inter-hemispheric comparisons such as synchrony, coherence, and amplitude asymmetries possible.

Vertical and horizontal EOG was recorded independently on two differential channels by way of four referential channels on the LTEEG system. The vertical EOG electrodes were placed 1 cm above and below the eye. The horizontal EOG electrodes were placed close to the inner and outer canthi.

Ideally, it would have been desirable to have had separate ECG and EMG channels for artifact identification, but this was not done due to the limited number of channels available and a priority given to additional EEG channels.

6.2.4 Performance measure

The task chosen to measure a subject's visuomotor performance in the new study had to meet several criteria. It had to be relatively undemanding and monotonous to increase the likelihood of a subject lapsing during the test, ideally requiring a continuous changing motor output. It also had to probe the subject's performance continuously so that there were no time periods when the instantaneous performance of the subject was unknown. The task software had to allow millisecond-scale synchronisation with external psychophysiological or other data collection procedures. Performance measures had to be sufficiently sensitive to allow differentiation between alert, drowsy, and absent performance. Finally, the task had to be suitable so that subjects could perform it for 1 hour or more.

A 1-D pursuit continuous tracking task (CTT) developed by Jones *et al*. (2000; 1993) was chosen for the study as it met all of the above criteria. This test is a component of a PC-based battery of tests (*SMCTests*™) developed for the quantitative assessment of upper-limb, sensory-motor, and cognitive functions. The use of a divided-attention task was decided against as it would likely increase arousal and performance (Stroh, 1971) and, hence, possibly lessen the likelihood of the occurrence of lapses.

SMCTests™ is based around a Pentium PC and two high-resolution 17" colour monitors: one for the tracking display and one for use by the assessor for task control and analysis. The two monitors can be controlled simultaneously to produce extremely smooth dynamic colour graphics. The subject input device was a steering wheel. The program contains seven 1-D tracking tasks and two 2-D tracking tasks. The primary error measure used by *SMCTests*™ to evaluate tracking performance was the mean absolute error – the mean value of the absolute distance/error (AE) between the target and the response at each sample.

It was hypothesized that the CTT was sufficiently undemanding and monotonous to facilitate deficits associated with under-arousal and loss of vigilance and, hence, allow the study of vigilance decrements. The CTT duration was considerably longer than the ATC study task duration (60 min *vs*. 10 min) so as to increase the probability of the subject lapsing due to the increased time-on-task (Doran *et al*., 2001). The simplicity of the CTT had the added advantage that it had a reasonably brief learning-curve. Having expertise to fully customize the software to suit the requirements of the study was another advantage since it was developed in-house.

The subjects used a steering wheel (395 mm diameter, wheel-to-screen gain = 1.075 mm/degree) to control an arrow-shaped cursor located near the bottom of the screen. The eye-to-screen distance was 136 cm.

Subjects were provided with an 8-s preview of a pseudo-random target (bandwidth 0.164 Hz, period 128 s) which scrolled downwards at a rate of 21.8 mm/s. The target signal was generated by summing 21 sinusoids evenly spaced at 0.00781 Hz intervals and with random phases. Figure 6-1 shows a screenshot of the target waveform. The task required smooth movements over a 175-deg range of the steering wheel and measured a subject's ability to keep the point of the arrow on the moving target. The position of the steering wheel was sampled at 64 Hz using a potentiometer mounted on the shaft of the wheel.

6.2.5 Video measures

It has been suggested that an independent validation criterion, such as video monitoring of changes in facial features for lapses, should be utilized to provide an objective indicator of lapsing (Lal and Craig, 2001). Head and facial features were therefore recorded from an analogue video camera (Sony Handycam) positioned 1 m in front of the subject using a frame rate of 25 Hz. The 'night mode' function of the camera was utilized to obtain a clear picture during low-light experimental conditions. The video was time-stamped using time-information provided by the LTEEG recording system.

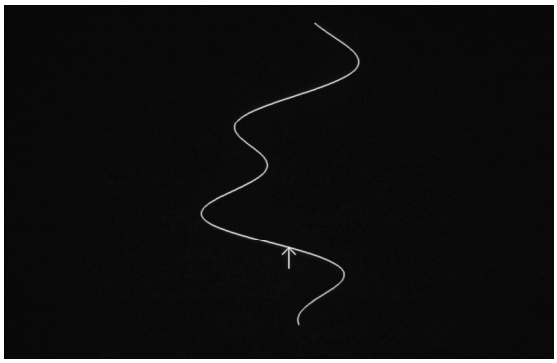

Figure 6-1 A screenshot depicting the random-preview target waveform used for the CTT. Subjects were instructed to keep the arrow-head of the cursor (which could only move horizontally) on the target waveform as closely as possible.

The availability of time-synchronized video provided an independent measure of alertness and enabled the researchers to confirm subjects' performance lapses. It could also be used to find the causes of movement artifacts in the physiological data.

6.2.6 Protocol

It has been suggested that subjects sitting in a comfortable chair in a quiet, dimly-lit, warm chamber who refrained from making unnecessary movements tended to lose vigilance and produce sporadic performance lapses (Makeig and Jung, 1996; Valley and Broughton, 1983; Van Orden *et al.*, 2000). Therefore, the subjects were instructed to try and remain as still as possible while performing the task. This also helped to minimize artifacts in the EEG and EOG.

Participants were instructed to refrain from taking any alertness-altering drug or medication (e.g., stimulants – coffee, amphetamine, tea; depressants – alcohol) 4 hours prior to all test sessions.

Subjects were encouraged to attend the test sessions after having a substantial lunch as this tends to make the subjects drowsier. All experimental sessions were held following lunch, between 12.30 p.m. and 5.00 p.m.

EOG was recorded while the subject performed sample eye movements prior to and following the tracking task. These included eye blinks, horizontal and vertical eye movements, and eye opening and closure. These sample eye movements were recorded so that they could serve as templates later to distinguish various eye movements from the EOG and EEG.

Subjects were instructed to strive to perform to the best of their ability and to keep their eyes open at all times during the tracking task. They were encouraged to go to the toilet prior to

the experiment as it was not possible to pause the recording once tracking had commenced. The investigator also ensured that distractions such as mobile phones were turned-off and watches were removed prior to the commencement of the tracking task. The room temperature was maintained at 22–25 ˚C.

There was a small possibility that participants might actually fall asleep at the wheel and (a) slump forward onto the wheel or (b) fall sideways from the chair. This was however considered unlikely, as the tendency is to retain posture despite lapsing and, at most, drop their head forward during a lapse. The principal investigator was on the lookout (state of readiness) to minimize the effects of such extended lapses by speaking and providing a warning to the subjects before they slumped too far forward.

The subject was questioned prior to the test session to determine the number of hours of sleep they had had the previous night and also the time they woke up. After the test, the investigator questioned the subject to determine how drowsy they became during the tracking and to find out whether they felt they had lapsed during any part of the test.

6.2.7 Test Procedure

Electrode locations were marked on the scalp by neurophysiology technicians according to the international 10-20 system layout. The scalp surface was then prepared and recording electrodes (gold or Ag/AgCl) were attached to the subject. This was followed by testing the EEG recording system to ensure that EEG and EOG signals were correctly displayed on the LTEEG PC. The technicians also ensured that the electrode impedances were less than 10 kΩ.

After confirmation of correct electrode attachment, sample eye movements were recorded from the subject for a period of approximately 2 min. The subject was instructed to perform the following eye movements at specific time intervals by the investigator in the following order:

1. 10 eye blinks
2. Horizontal eye movements: centre (C), left (L), C, right (R), C, L, C, R, C
3. Vertical eye movements: C, up (U), C, down (D), C, U, C, D, C
4. Eyes closed (EC), eyes open (EO), EC, EO
5. 5 blinks

Each left, right, up, and down movement was calibrated to be a 10-deg deviation from the centre position. EO and EC were each 5 s in duration. The vertical and horizontal eye movements were spaced apart by 1–2 s.

Following the eye movement recording, the subject performed the CTT with EEG, EOG, and video recording for a period of 1 hour. They were instructed to stay alert and perform the task to the best of their ability. Figure 6-2 shows a subject being prepared for the CTT study.

At the end of the tracking task, each subject was asked how drowsy they felt during the tracking and whether they considered they had "lapsed" at any point.

Eye movement recording was repeated in case of electrode movement during the tracking task.

Following the second eye movement recording, the EEG and EOG electrodes were removed from the subject and their visual acuity measured using a Snellen chart (3 m - USL Medical, Auckland, New Zealand).

Each subject attended two sessions held at least one week apart (mean 17 days, range 7–50 days).

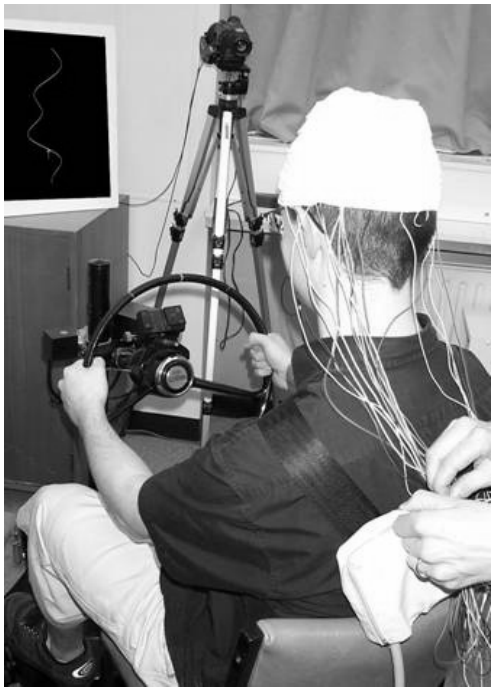

Figure 6-2 A photo depicting a subject being prepared for the CTT tracking task. The EEG and EOG electrodes have been attached to the scalp after preparation and a bandage has been wound around the head to ensure minimal movement of electrodes during the test. The photo also shows the video camera which recorded head movement and facial features, and the steering wheel the subjects used to control the cursor which appeared on the screen in front of them.

CHAPTER 7

Characteristics of Lapses

7.1 Introduction

An important outcome of the study outlined in Chapter 6 was in characterizing lapses. These characteristics are the subject of this chapter. Before providing these findings however, the state-of-art is briefly reviewed.

Even well rested individuals who are not sleep-deprived may experience 'sleep-related states' without a preceding phase of subjectively experienced drowsiness (Sagberg, 1999). Drowsy individuals performing an extended active task, such as driving, often cycle rapidly between periods of wake and sleep, as exhibited by cyclical variation in both EEG power spectra and task performance measures (Makeig *et al.*, 2000). Although well documented in sleep-restricted people, these episodes are less well described in rested non-sleep-deprived individuals.

Lapses of responsiveness ('lapses') are brief episodes in which a subject unintentionally stops responding to the task they are performing. By contrast, the related term *microsleep* is usually used to describe brief episodes (min 1–15 s, max = 14–30 s) of EEG–defined sleep (Harrison and Horne, 1996; Hemmeter *et al.*, 1998; Priest *et al.*, 2001; Tirunahari *et al.*, 2003; Valley and Broughton, 1983). While many lapses are associated with EEG-defined microsleeps, and hence are apparently caused by low arousal, the link between EEG defined sleep and lapse behaviour is generally not strong (Ogilvie, 2001). Non-arousal-related lapses in sustained or task directed attention also occur and may show quite distinct characteristics (Neale *et al.*, 2005; Parasuraman and Davies, 1984).

Numerous groups have demonstrated lapses of performance under monotonous task conditions with and, less commonly, without sleep deprivation. A number of studies have

reported lapses in sleep-deprived subjects performing a variety of auditory and visual sustained attention tasks, and mental tasks (Cajochen *et al.*, 1999; Dinges *et al.*, 1997; Doran *et al.*, 2001; Kleitman, 1963; Schroeder *et al.*, 1994; Torsvall and Åkerstedt, 1988). In addition, lapses have been reported during simulated night-time driving (Arnedt *et al.*, 2005). Bills (1931) reported an increase in the number of 'blocks' (or lapses) in non-sleep-deprived subjects with time-on-task in subjects performing a mental task. More recently, Van Orden *et al*. (2000) reported lapses in a group of non-sleep-deprived subjects performing a continuous 2D tracking task during normal waking hours.

Sleep deprivation has a strong influence on level of drowsiness and, consequently, task performance. Using a 10-min psychomotor vigilance task (PVT), Dinges *et al*. (1997) determined that the maximum lapse rate after 7 days of sleep restriction (5 hours of sleep per night) in a group of 20 young adults was 24 lapses/h (reaction time >500 ms). Lapsing did not increase between the second baseline night and the first night of sleep restriction but increased after that. After 70 hours of total sleep deprivation, Doran *et al*. (2001) measured a maximum PVT error rate (false start trials and anticipatory responses <100 ms) of 96 errors per hour. In a study by Torsvall and Åkerstedt (1988), participants 'dozed off' 7.4 times during a 45-min discrete visual vigilance task which commenced at 01:30 hours. Doran *et al*. (2001) suggested that performance during sleep deprivation is highly unstable because of the subjects rapidly fluctuating between states which cannot be defined as fully awake or asleep because of the influence of sleep-initiating-mechanisms.

Time-on-task also plays a key role in performance fluctuations during mental tasks (Bills, 1931; Dinges *et al.*, 1997; Thiffault and Bergeron, 2003). Mackworth (1969) suggested that decline in performance during a monotonous task may be because of habituation of the neural response because of repetitive stimuli.

While circadian factors have a strong effect, task and environmental conditions also influence task performance. Almost all studies examining driver fatigue suggest that it occurs more rapidly on monotonous roads such as highways (Thiffault and Bergeron, 2003). Subjects sitting in a comfortable chair in a quiet, dimly-lit, and warm room, who have been asked to refrain from making unnecessary movements, tend to lose vigilance and have intermittent lapses in task performance, even when well-rested (Makeig and Jung, 1996; Valley and Broughton, 1983; Van Orden *et al.*, 2000).

In the study described in Chapter 6, lapses of responsiveness were observed in non-sleep-deprived young adults during normal working hours. A conservatively defined sub-category of lapses during which clear behavioural signs of sleep, such as eyelid closure and head-nodding were evident (i.e., BM) was also investigated.

The focus of this chapter is on describing the occurrence and characteristics of lapses including BMs in a group of subjects superficially considered unlikely to fall asleep during a task. Conservative estimates of BM rate and duration were derived by searching for clear lapses of task responsiveness in tracking performance (indicated by the cessation of movement of the response cursor for an extended period while the target is moving – henceforth referred to as a *flat spot*) occurring concurrently with behavioural sleep observed on the video recordings (*video BM*). In addition to the conservative estimate, rate and duration estimates for lapses based on the occurrence of either a flat spot in the tracking response or a video BM were calculated. Finally, changes in EEG spectral power in the standard bands associated with lapses were investigated.

7.2 Performance and EEG analysis

7.2.1 Flat spot detection

Lapses in tracking performance are most obvious when the response cursor simply stops moving for an extended period while the target is moving or when the tracking response is non-coherent with the target. Only the first category (flat spots), were included in an intentionally conservative analysis, as lapses in the second category are difficult to identify with confidence. Flat spots occurring when the target velocity is approximately zero (at turning points) were not counted, as at these times the subject can track adequately without moving the response cursor.

A schematic of the procedure used to detect flat spots is shown in Figure 7-1. The target and response signals were low-pass filtered with a cut-off at 5 Hz using an eighth-order bi-directional Butterworth filter. Target and response sections were then marked using the following criteria: (a) *flat target region* − a flat segment of the target, defined as a section of at least 300 ms duration, within which the deviation was ≤1.54 mm and (b) *flat response region* − a flat segment of the response, defined as a section of at least 1500 ms duration, within which the deviation was ≤0.77 mm. A 2.0-s 'start zone' and a 1.25-s 'end zone' were marked within each flat response region.

A flat response region was classified as a flat spot if any of the following criteria were satisfied:

a) there were no flat targets within the 'start' or 'end' zones of the flat response

b) only one flat target occurred within the 'start' or 'end' zones of the flat response

c) the RMS error between the target and the response during the flat response region was >15.0 mm

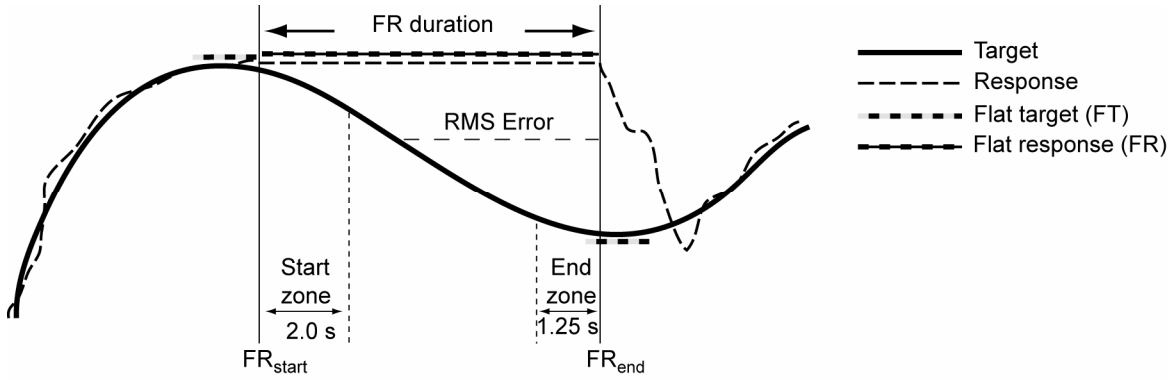

Figure 7-1 Flat-spots were detected using flat segments of the target and response waveforms, RMS error between the target and the response during a flat response (FR), and the duration of a flat response (FR duration). Flat target (FT) sections are segments in the target waveform of at least 300 ms duration and with a within-segment deviation less than 1.5 mm. The 2.0 s start zone and the 1.25 s end zone are also marked within the FR (a FT can occur within either zone but not both).

d) the duration of the flat response region was ≥6.0 s.

Criterion 'c' was necessary for cases where a flat response coincided with two or more flat target regions, as such a flat response would not be detected as a flat spot using criterion 'a' or 'b'. However, by utilizing the large error between the target and response during a flat response, it was possible to correctly identify many of these as flat spots. Visual inspection of the tracking data confirmed that an RMS error threshold of 15.0 mm was sufficiently sensitive to detect these types of flat spots without introducing false positives. The duration of the longest contiguous flat target region was 5.0 s, so any flat responses ≥6.0 s had to be flat spots.

7.2.2 Video rating

The video recording of each session was conservatively rated by the author, without knowledge of the corresponding tracking performance. Being 'blind' to tracking performance ensured that an independent measure of alertness was obtained from the video data. The video was rated on a 6-level scale: 1 = alert, 2 = distracted, 3 = forced eye closure while alert, 4 = light drowsy, 5 = deep drowsy, and 6 = sleep (including microsleep). Criteria similar to Weirwille *et al*. (1994) were used to define the video rating scale. Alert periods were identified by the presence of features such as fast eye blinks and normal facial tone in video recordings. Intervals when the subject appeared diverted from the task at hand were rated as 'distracted'. Instances of the subject intentionally closing their eyes while alert (typically to relieve eye fatigue) were marked as 'forced eye closure'. Instances of the subject appearing subdued, with slower blinks, were rated as 'light drowsy'. 'Deep drowsy' was identified by a paucity of eye movements, decrease of facial tone, and partial closure of the eye lids. Based on video alone, deciding between light and deep drowsy was found to be

particularly difficult. Sleep events (i.e., video BMs) were identified by prolonged eye-lid closure, sometimes accompanied by rolling upward or sideways movements of the eyes, head-nodding, and often terminated by waking head jerks. Transitions in the video recording had a time resolution of 1.0 s.

7.2.3 Definite BM rate and duration

Intervals in which flat spots and video BMs overlapped in time were defined as *definite BMs*. To ensure that the estimated event rates and durations were conservative, slightly different methods were used to calculate each. Definite BM duration was obtained by finding all samples for which a video BM and a flat spot occurred simultaneously (Figure 7-2). To obtain a conservative estimate of definite BM rate, a logical 'OR' was applied between video BMs and flat spots. This ensured that only a single event was counted if there were multiple flat spots during a video BM and vice versa. Except when calculating rate, the duration method was always used to identify definite BMs.

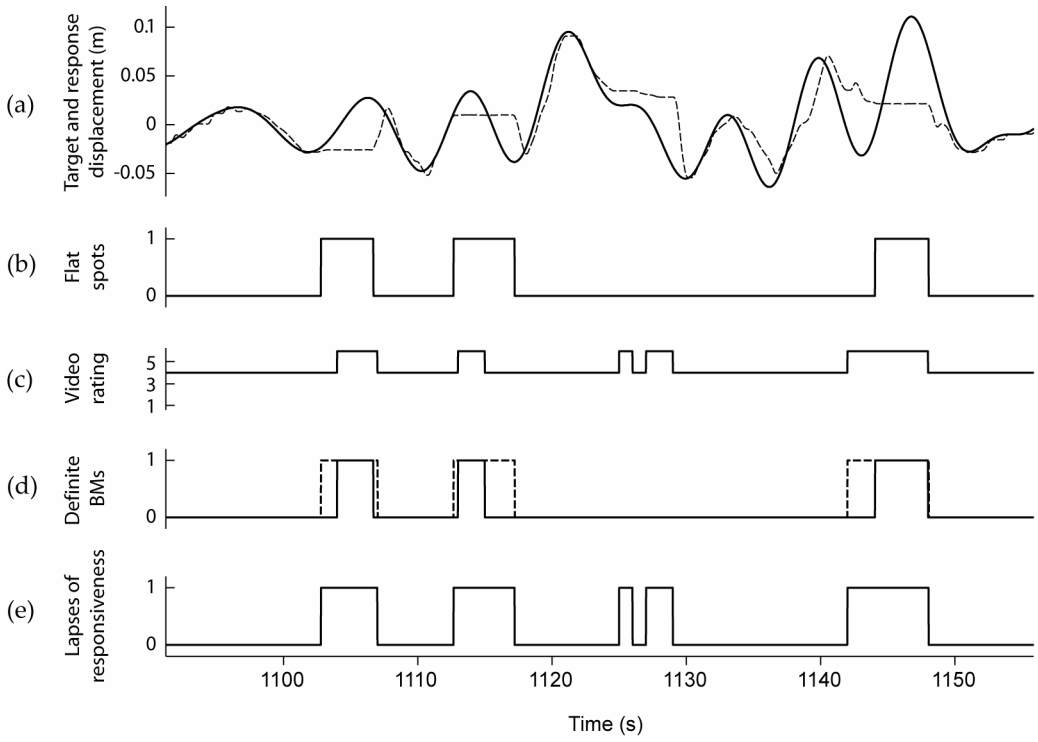

Figure 7-2 An illustrative example of the methodology used to obtain estimates of definite BM and lapse counts and durations. Data from a subject approximately 18 min into the 1-h tracking task: (a) tracking behaviour including target (solid) and response (dotted), (b) detected flat spots, (c) independent video rating of the subject's level of alertness, (d) definite BMs in terms of lower bounds for count (dashed) and duration (solid), and (e) lapses. There is little doubt that a lapse occurred between 1124 and 1129 s but, as no response flat spot was identified, no definite BM was identified. However, according to the video rating, two 'sleep' events occurred during the same interval and hence two lapses were identified.

As the pseudo-random target had a period of 128 s, a bin-width of 128 s was used to calculate event rate changes as this provided target segments of identical difficulty and enabled comparison of performance across epochs. Event rate was calculated in terms of the onset of events.

7.2.4 Lapse rate and duration

The presence of either a video BM *or* a flat spot (cf. *both* needed for a definite BM) provides a sound, if conservative, indicator of the presence of a lapse. The minimum detection duration for all events (flat spots, video BMs, definite BMs, and lapses) was 1.0 s. Figure 7-2 illustrates how the lapse and definite BM rate and duration measures were derived.

7.2.5 EEG analysis

A notch filter was applied to remove 50 Hz interference from the EEG and independent components analysis was applied to remove eye blink artifacts (Delorme and Makeig, 2004; Jung *et al.*, 2000b). A window size of 512 samples (2 s) and an overlap of 50% between successive windows were used to calculate power spectra via a 40^{th}-order autoregressive Burg model; a high model order was required to obtain adequate separation of the spectral bands of interest. Data in each window was de-trended prior to the calculation of power spectra. A window overlap of 1 s allowed a temporal resolution of 1 s for the spectral power in the delta (1.0–4.5 Hz), theta (4.5–8.0 Hz), alpha (8.0–2.5 Hz), beta (12.5–25.0 Hz), gamma (25.0–45.0 Hz), and high (45.0–100 Hz) bands. To remove electrode-pop artifacts, each derivation was normalized into z-scores. Epochs containing large EEG artifacts (absolute z-score >30) were rejected and excluded from further analysis.

Each of the four behavioural metrics (flat spots, video BMs, definite BMs, and lapses) were decimated to a resolution of 1 Hz and provided a proxy for differentiation between the states of not-lapsing (0) and lapsing (1). The mean and standard deviation of the EEG power for a particular band during the non-lapsing samples was calculated for each of the behavioural metrics. These values were then used to transform EEG power in each band during lapses to spectral z-scores and averaged across all lapse samples to obtain a mean z-score for each band. An overall mean z-score for each band for each subject was calculated by averaging over all channels and across both sessions.

7.2.6 Statistical analysis

Unless otherwise stated, paired *t*-tests were used for statistical tests. Mixed-model ANOVAs with repeated-measures were used to investigate differences in lapse metrics and EEG power. The data were tested for *sphericity* and, if violated, significance values were adjusted

by the Greenhouse-Geisser correction (Greenhouse and Geisser, 1959). Sphericity assumes that repeated-measures are uncorrelated and that there is equality of variances of the differences between levels of the repeated measures factor (Thomas *et al.*, 2005). The rate and duration of flat spots, video BMs, and tracking error were separately analysed with Session (1 vs. 2) and Epoch (1–28) as within-subjects factors and Subject as the random factor.

7.3 Results

7.3.1 Basic characteristics of lapses

Based on the conservative definite BM rate estimate, eight of the 15 subjects had one or more definite BMs during their two 1-h sessions. Six subjects had 30 or more definite BMs and two subjects had 123 and 144 definite BMs respectively over the two 1-hr sessions. Table 7-1 gives a summary of mean, standard error, and ranges of event rate and mean duration of flat spots, video BMs, definite BMs, and lapses for all subjects, and for the eight subjects who had at least one definite BM. For subjects who had one or more definite BMs, the first occurred 22.3 ± 3.4 min into the session (mean ± SE, range 5.5–51.3).

The rates and mean durations of flat spots, video BMs, definite BMs, and lapses for individual subjects are shown in Figure 7-3. This shows considerable inter-subject variability in the rate of events and, to a lesser extent, in the duration of events. Figure 7-4 shows a histogram of the distribution of the duration of lapses and definite BMs. This indicates that the most frequent definite BMs are of 2–3 s duration and that the rate of longer events decreases approximately exponentially reaching a plateau of less than 1 per hour for durations longer than approximately 9 s.

Over all 15 subjects, 2.0 ± 1.0 % (range 0.0 – 12.6) of each session was classified as definite BM and 5.0 ± 2.0 % (range 0.0–26.3) in the state of lapsing.

There was no correlation between lapse rate and age (r = -0.012, t-test p = 0.97), nor between the proportion of the task spent lapsing and age (r = -0.05, t-test p = 0.86). This suggests that propensity to lapse is not related to age, at least within the range of ages studied. However, this does not mean that a link between the incidence of lapsing and age would not be found if a wider range of ages was incorporated.

There was no difference between sessions 1 and 2 in terms of the number of subjects who had a definite BM (5 *vs*. 8, McNemar's test p = 0.25) and only a marginal increase in the rate of definite BMs (10.8 *vs*. 19.6 /h, p = 0.085).

Table 7-1 Mean, standard error, maximum, and minimum values of event rate and duration of flat spots, video BMs, definite BMs, and lapses. Results are shown for all subjects (*N* = 15) and for those who had definite BMs (N = 8). As all subjects did not have flat spots or video BMs, the mean event duration was calculated over different *N* for each type of event.

	All subjects mean ± SE (min,max)	Definite BM subjects mean ± SE (min,max)
Event rate (/h)		
Flat spots	23.3 ± 8.9 (0.0, 102.5)	43.3±13.2 (9.5, 102.5)
Video BMs	35.1 ± 12.2 (0.0, 142.0)	65.1 ± 16.8 (8.5, 142.0)
Definite BMs	15.2 ± 5.9 (0.0, 72.0)	28.5 ± 8.8 (2.0, 72.0)
Lapses	39.3 ± 12.9 (0.0, 141.5)	72.5 ± 16.9 (16.0, 141.5)
Event duration (s)		
Flat spots	3.4 ± 0.4 (1.9, 5.4); N = 10	3.8 ± 0.4 (2.1, 5.4)
Video BMs	3.4 ± 0.5 (1.0, 7.7); N = 12	4.0 ± 0.7 (1.9, 7.7)
Definite BMs	3.2 ± 0.4 (1.8, 4.6); N = 8	3.2 ± 0.4 (1.8, 4.6)
Lapses	3.4 ± 0.5 (1.0, 8.3); N = 14	4.4 ± 0.7 (2.0, 8.3)

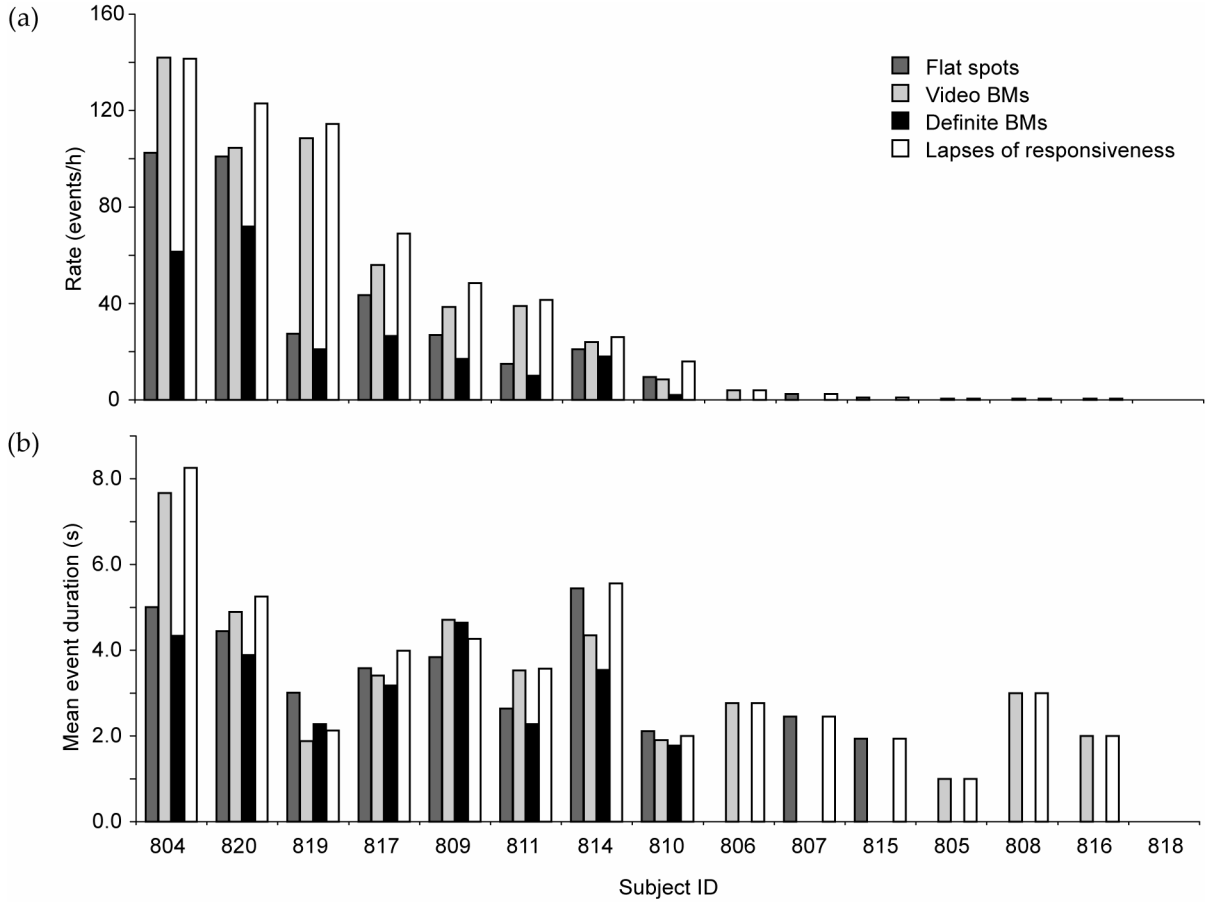

Figure 7-3 Inter-subject variability: (a) rates and (b) mean durations of flat spots, video BMs, definite BMs, and overall lapses during the tracking task (over two 1-h sessions). Subjects are ordered in descending order of lapse rate.

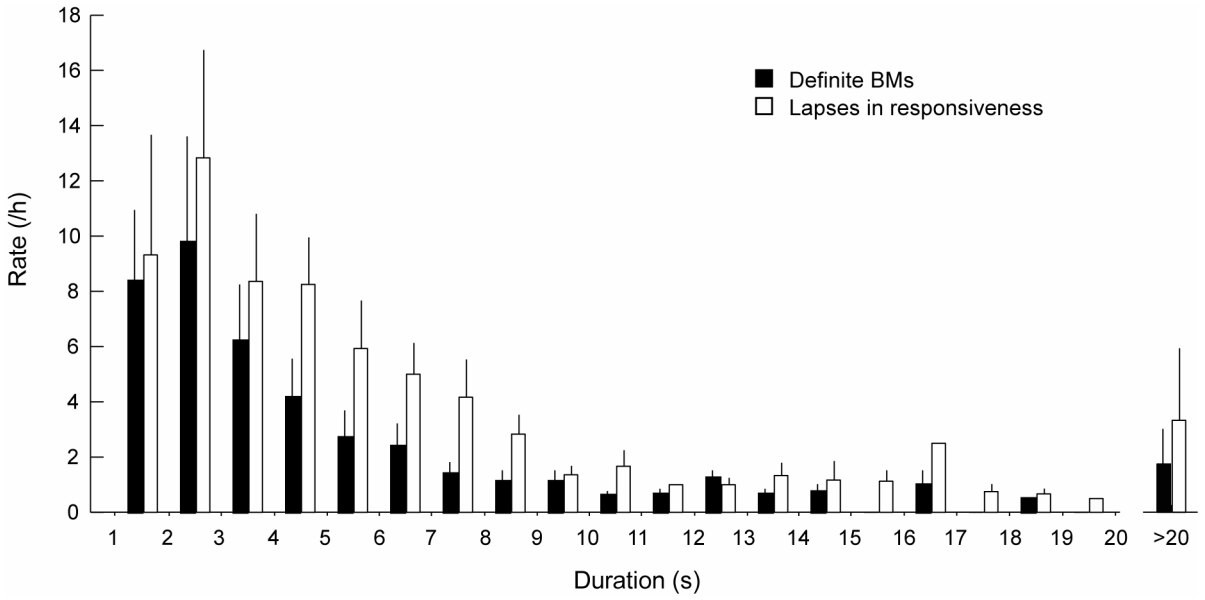

Figure 7-4 A histogram showing the distribution of the durations (mean ± SE) of definite BMs (filled bars) and lapses (unfilled bars). The bin width is 1.0 s. The final pair of bars depict lapses >20 s duration.

The rate of flat spots without a coincident video BM was 5.0 ± 1.6 events/h, whereas the rate of video BMs without a coincident flat spot was 19.1 ± 7.1 events/h. Visual analysis of the tracking and video data (e.g., Figure 7-2) indicated that this difference is primarily due to the conservative approach adopted to define flat spots; many of these video BMs coincided with a slowly drifting, non-coherent tracking response. There was no difference found between flat spot and video BM event durations ($p = 0.540$). However, a marginal difference was found between the event rates ($p = 0.058$).

After completing the tracking, subjects were asked whether they were aware of having any "lapses" and/or sleep episodes during the session. All but one of the 14 subjects who had at least one lapse in a session considered that they had lapsed during that session. Conversely, 4 subjects considered they had lapsed during a session in which no lapses were evident.

From the video rating, samples were classified as (mean ± SE): alert = 57.5 ± 8.4 %, distracted = 0.2 ± 0.1 %, forced eye closure = 0.3 ± 0.1 %, light drowsy = 28.6 ± 6.0 %, deep drowsy = 9.1 ± 3.3 %, and video BM = 4.2 ± 1.7 %. Flat spot rates (mean ± SE) during each video rating level were: alert = 0.8 ± 0.4 /h, distracted = 130.6 ± 125.8 /h, forced eye closure = 60.5 ± 40.2 /h, light drowsy = 21.6 ± 7.0 /h, deep drowsy = 67.3 ± 20.0 /h, and video BM = 179.3 ± 27.4 /h. Repeated-measures ANOVA showed a strong main effect of video-rating level on flat spot rate ($F = 21.007$, $p < 0.0001$).

7.3.2 Time course of lapses

Figure 7-5 shows the time course of the rate and duration of flat spots and video BMs for both sessions of all subjects (N=15) over 128-s epochs. As expected, a gradual increase in the flat spot and video BM rates was observed with time-on-task until the middle of the session. Conversely, the monotonic decrease in the event rates in the latter half of both sessions was unexpected. Event durations were found to be relatively stable throughout the sessions. Tracking error, defined as the absolute difference between the target and user response, and averaged over the 128-s equal-difficulty target epochs, is shown in Figure 7-5 (c). This shows a similar time course to the flat spots and video BMs.

The sphericity assumption was not met and, hence, Greenhouse-Geisser correction was applied to the mixed-model ANOVA with repeated-measures. A main effect of Session was only observed in the flat spot rate ($F = 4.621$, $p = 0.050$). Tracking error varied with Epoch (main effect; $F = 4.102$, $p = 0.022$) and showed a quadratic trend ($F = 7.550$, $p = 0.016$). Similarly, the video BM rate also varied with Epoch (main effect, $F = 2.810$, $p = 0.050$) and showed a quadratic trend ($F = 8.326$, $p = 0.012$), confirming the apparent improvement in performance toward the end of the session. There was no interaction between Session and Epoch for any measure.

7.3.3 EEG changes during lapses

Figure 7-6 shows the mean z-scores for power in each EEG spectral band during flat spots, video BMs, definite BMs, and lapses, relative to non-lapsing. It indicates a mean increase in power with lapses in the delta, theta, and alpha bands, and a decrease in the beta, gamma, and high-frequency bands. Mixed-model repeated-measures ANOVA, with Metric (flat spots, video BMs, definite BMs, and lapses) and Band (delta, theta, alpha, beta, gamma, and high) as factors, Subject as random factor, and power z-score as the dependent variable, showed a main effect of Band ($F = 9.847$, $p = 0.002$) but not Metric ($F = 0.506$, $p = 0.581$). Within-subjects contrasts showed linear ($F = 25.455$, $p = 0.001$) and cubic ($F = 7.664$, $p = 0.028$) trends for Band. There was no interaction between Metric and Band.

The mean power z-score in each EEG spectral band for flat spots that did not overlap with video BMs was also calculated. A scaled-down version of the pattern for all flat spots shown in Figure 7-6 was observed, with a less positive mean z-score for delta, theta and alpha bands and a less negative mean z-score for beta, gamma, and high-frequency bands (Figure 7-7). A separate mixed-model ANOVA with repeated-measures showed that there was a main effect of flat spot Type (all flat spots *vs*. flat spots without video BMs, $F = 5.116$, $p = 0.050$) and an interaction between Type and Band ($F = 4.991$, $p = 0.013$). There was a linear trend across band ($F = 9.848$, $p = 0.012$) and also an interaction in the linear trend between Flat Spot Type

and Band ($F = 9.087$, $p = 0.015$), with the gradient of the trend for flat spots without video BMs being 53% smaller than that for all flat spots. This is also confirmed by observing that the power *z*-scores for flat spots without video BMs (Figure 7-7, right) were scaled down by approximately half when compared to the power z-scores all flat spots (Figure 7-7, left).

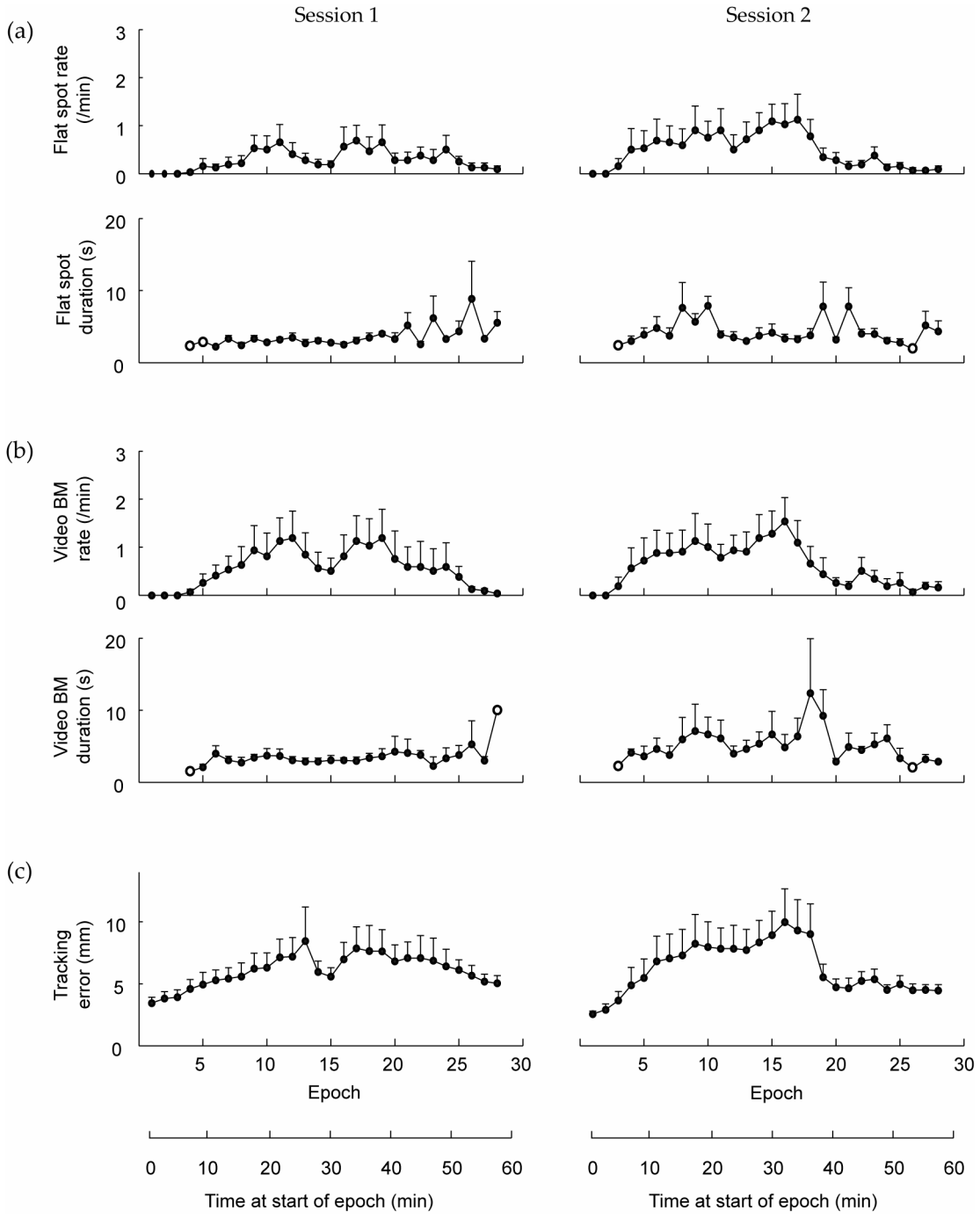

Figure 7-5 Time course of performance metrics (mean ± SE) used to determine lapse characteristics for all subjects for sessions 1 and 2: (a) flat spot rate and duration, (b) video BM rate and duration, and (c) tracking error. The epoch size was 128 s corresponding to a cycle of the periodic pseudo-random target. A hollow circle indicates epochs for which SE could not be calculated as only one subject contributed to the mean.

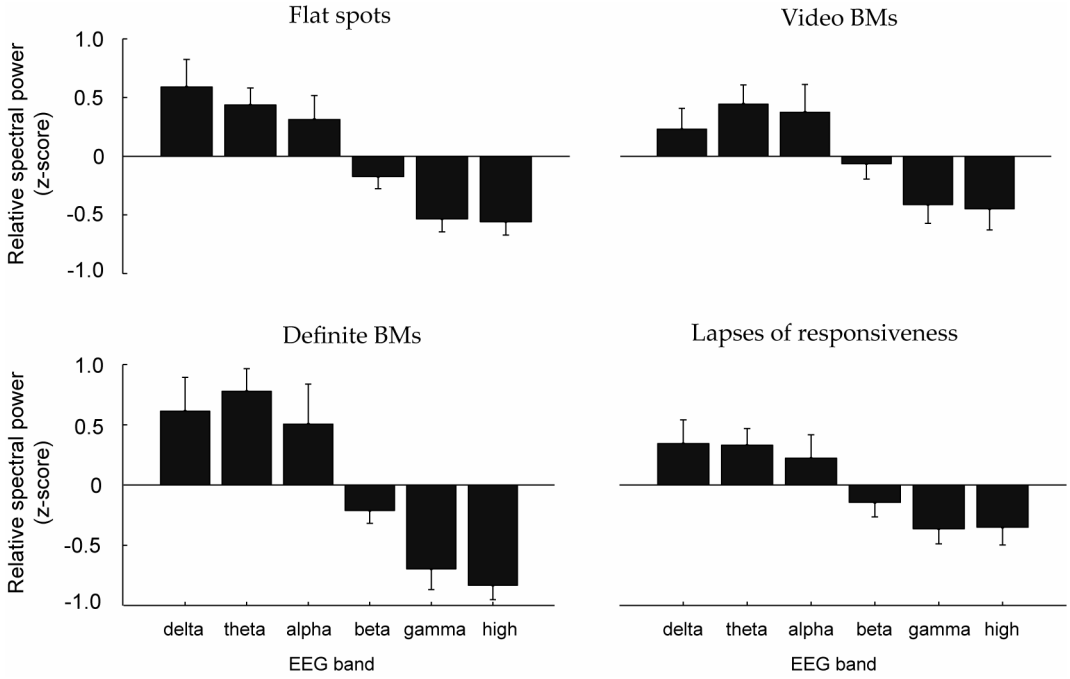

Figure 7-6 Changes in EEG spectral power during lapses in terms of *z*-scores (mean ± SE) relative to power during the non-lapsing state averaged across all channels, both sessions, and for all subjects during flat spots, video BMs, definite BMs, and lapses of responsiveness.

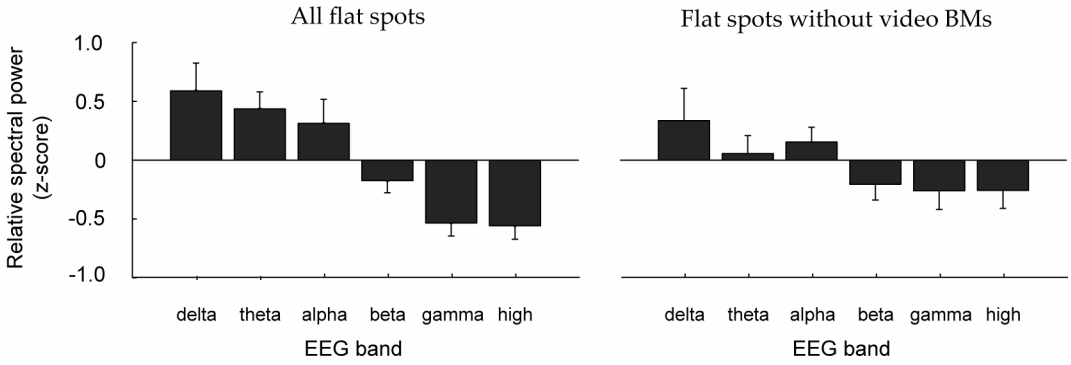

Figure 7-7 Changes in EEG spectral power during flat spots in terms of *z*-scores (mean ± SE) relative to power during the non-lapsing state averaged across all channels, both sessions, and for all subjects during all flat spots and flat spots without video BMs.

EEG power showed a weak correlation with lapses in the delta ($r = +0.06 \pm 0.03$; *t*-test, $p = 0.040$), theta ($r = 0.06 \pm 0.03$, $p = 0.009$), gamma ($r = -0.07 \pm 0.02$, $p = 0.006$), and high ($r = -0.09 \pm 0.03$, $p = 0.004$) bands. Similar correlations were obtained for definite BMs: delta ($r = +0.09 \pm 0.04$; *t*-test, $p = 0.04$), theta ($r = +0.13 \pm 0.04$; $p = 0.007$), gamma ($r = -0.11 \pm 0.02$; $p = 0.003$), and high ($r = -0.13 \pm 0.02$; $p < 0.001$). However, the mean absolute correlation across bands between lapses and definite BMs was not different (0.07±0.02 *vs*. 0.10±0.01; $p = 0.224$).

7.4 Discussion

This is the first study to have investigated the characteristics of complete lapses of responsiveness during an extended continuous tracking task in non-sleep-deprived subjects. It demonstrates the occurrence of frequent unequivocal lapses in a substantial proportion of normal non-sleep-deprived subjects during normal working hours. The study has also shown increases in both lapsing and tracking error during the first 30 min of a 1-hour session but subsequent decreases over the second 30 min, despite the absence of external temporal cues.

The high incidence of lapsing in young, healthy, and non-sleep-deprived adults carrying out a continuous visuomotor task during normal work hours was unexpected. While the experimental conditions were intentionally conducive to sleep (i.e., warm room, lights down, quiet environment, post-lunch, circadian low), the requirement of a continuously changing motor response, being recorded on video, being in an overt experiment, having been instructed to attend to the tracking task as accurately as possible at all times, and not being sleep-deprived, could well have prevented most, if not all, lapses. If these factors did have an inhibitory effect, it certainly was not sufficient to prevent lapsing in nearly all of the subjects.

Most previous studies have used sleep deprivation as a means of inducing and/or increasing the likelihood of a subject lapsing (Cajochen *et al.*, 1999; Dinges *et al.*, 1997; Doran *et al.*, 2001; Gillberg and Åkerstedt, 1998; Roge *et al.*, 2003). However, Van Orden *et al*. (2000) reported that 9 out of 15 non-sleep-deprived participants had performance lapses on a 53-min 2D compensatory tracking task, although the rate and characteristics of these lapses were not reported.

The rate of video BMs was considerably higher than that for definite BMs (1.97 times), which appears to be primarily because of BMs causing events other than flat spots in tracking, such as the response cursor drifting away from the target. Conversely, some flat spots occurred without coincident behavioural signs on video, which appears to reflect a combination of the conservative approach taken to rating the video (e.g., questionable lapses were rated deep drowsy) plus an absence of sleep-related signs with some lapses of performance. It is unclear whether these latter lapses are arousal-related microsleeps or non-arousal-related lapses of sustained attention or diverted attention, although cases of overt diverted attention were eliminated as these were identified in the video rating.

Nearly all subjects who had one or more lapses in responsiveness were aware that they had lapsed during the session. This agrees with anecdotal evidence of drivers being aware that

they have briefly 'nodded off' while driving at night. However, four subjects considered they had lapsed during one or more sessions in which no lapses were evident. This may indicate that some subjects have a tendency to over-estimate lapses but could also reflect non-arousal-related lapses in attention.

On both sessions, the mean absolute tracking error and rates of flat spots and video lapses were all seen to increase progressively up to around the middle of the session but then decrease over the second half. This improvement in performance was surprising considering there was no break in the task, and subjects were not explicitly aware of the remaining task time. However, humans have been shown to possess an ability to estimate time on a 1-h scale (Aschoff, 1985; Aschoff, 1998). Therefore, one can speculate that subjects may have been able to gauge the approximate time remaining and to suppress or reverse the effects of drowsiness in anticipation of the end of the test. A similar trend was observed by Van Orden *et al*. (2000) where the mean tracking error of subjects performing a 2D compensatory tracking task increased until approximately 12 min into the task, followed by a plateau or possibly a decrease until the end of the 53-min test. It was shown definitively that performance improved towards the end of the 1-hr tracking session, despite the absence of external temporal cues. Another intriguing possibility is that lapses during a task act as mini 'rest periods' for the subject and that the cumulative effect of such lapses may have a restorative effect, leading to reduced drowsiness and improved performance.

The importance of using a substantial duration (1 hour) for the tracking task is demonstrated by the onset time of the first definite BM in a session which ranged from 5.5 to 51.3 min. The mean time of 22.3 min for the occurrence of the first definite BM is comparable with that seen by Thiffault and Bergeron (2003) in a study of driver fatigue in which they found that the impact of fatigue is robust and appeared quite early during each driving session, with a marked peak occurring after 20 to 25 min of driving. Shorter tests of sustained vigilance, such as the PVT, typically of 10-min duration, might fail to induce BMs in subjects who are not sleep-deprived because of the relatively short duration of time the subject is required to be attentive. Other studies have suggested that signs of fatigue may only be observed in some subjects after about 60 min of driving (Skipper and Wierwille, 1986) or on a vigilance task (Galinsky *et al.*, 1993).

It needs to be emphasized that the level of lapsing seen in the current study is unlikely to be paralleled in the somewhat equivalent on-road task of driving a vehicle. Fatigue and boredom are likely to be more evident and more difficult to counter in a monotonous task such as the CTT as opposed to on-road driving. Degraded performance and level of sleepiness in such real-life situations are likely to be substantially less because of the higher level of stimulation (Åkerstedt *et al.*, 2005) and the far greater consequences of lapses. This

notwithstanding, numerous studies have demonstrated a convincing link between tiredness, fatigue, and falling asleep and fatal accidents on the road (Cummings *et al.*, 2001; Horne and Reyner, 1995b; Lal and Craig, 2001; Sagberg, 1999; Stutts *et al.*, 2003).

The presence of BMs was determined conservatively by a requirement for coincident and independent video-based BMs and non-response tracking. Thus, irrespective of uncertainties about underlying mechanisms of microsleeps in terms of phasic changes in arousal and/or attentional systems in the brain, their frequent occurrence in normal non-sleep-deprived subjects on sustained continuous tasks is indisputable. Although eye fatigue is likely during an extended tracking task, as are consequent relief measures such as temporary closing of eyes, these were not considered a confounding factor in causing BMs. The video rating scale had a category of 'forced eye closure' to take such events into consideration and to differentiate them from potential video BMs. It was also noted that forced eye closures of even a few seconds do not cause deterioration in tracking performance as subjects are able to plan and execute predictive movements to maintain the cursor on a low-bandwidth preview target.

Changes in EEG power with poorer performance and with reduced levels of alertness have been seen previously (Huang *et al.*, 2001; Jung *et al.*, 1997; Makeig and Inlow, 1993; Makeig and Jung, 1995; Makeig and Jung, 1996; Makeig *et al.*, 1996; Makeig *et al.*, 2000; Santamaria and Chiappa, 1987). The EEG analysis of the current study extends this by showing significant changes in the EEG spectral bands during flat spots, video BMs, definite BMs, and lapses, compared with the non-lapse state. The changes involved increases in delta, theta, and alpha power and decreases in beta (small), gamma, and higher frequency power. Being most definitively related to level of arousal, it is not surprising that changes were most pronounced during definite BMs. However, the correlations between EEG band power and definite BMs were, at best, low. The bands showing the greatest mean correlation with definite BMs were theta, gamma and higher frequency activity. Theta power has been previously shown to correlate with auditory alertness (Huang *et al.*, 2001). Lal and Craig (2005) has reported slow wave activity (delta and theta) to be correlated with driver fatigue.

Similar trends were observed in EEG power when flat spots occurred without apparent video sleep, although less pronounced. Hence, these are likely to represent 'shallower' lapses. This seems particularly likely given the observation that lapses typically begin with flat spots, with video BM appearing slightly later. The trend may also reflect the conservative identification of video BMs.

Despite some initial apprehension as to whether subjects would lapse at all during the 1-h tracking task, and the conservative procedure used to identify both BMs and lapses, this

study has demonstrated that serious lapses can occur in young healthy non-sleep-deprived adult males to a much greater extent than previously recognized. This has major implications for occupations that require sustained alertness over long periods of time. From a safety point of view, it would clearly be very desirable to counter such lapses by early detection and wake-up systems. The next chapter describes the development and performance of a lapse detection system based upon subtle indicators of lapses in the EEG.

CHAPTER 8

Detection of Lapses from the EEG

8.1 Introduction

Chapter 7 demonstrated that lapses are a common phenomenon, even in non-sleep-deprived subjects performing a monotonous task during normal work hours, with subjects lapsing frequently (39.3 ± 12.9 /h). Analysis also revealed that these lapses are relatively brief (3.4 ± 0.5 s). As pointed out in § 1.2, the occurrence of such unintentional lapses in occupations which require sustained attention can lead to serious accidents due to persons failing to respond adequately (or at all) during a critical situation. Therefore, a device capable of monitoring a person in real-time and providing an alarm at the start of a lapse or, better still, a warning of an impending lapse would be invaluable. The first step in the development of such a device is to design a system capable of detecting brief lapses of responsiveness with good temporal resolution. The focus of this chapter is on evaluating various EEG cues and the effectiveness with which they can be used to detect lapses.

The work described in this chapter utilizes the EEG data collected from the continuous tracking task study described earlier (Chapter 6). EEG was recorded from 16 bipolar derivations for a period of one hour in each of two sessions while the subjects performed the continuous tracking task. This resulted in 2 hours of EEG data (16 derivations) per subject. EEG data of subjects who had at least one definite BM over the two sessions were selected for lapse detector design (N=8).

Chapter 7 revealed that lapses are associated with EEG spectral changes, albeit with a low correlation (mean absolute correlation 0.07 ± 0.02). Therefore, the initial work in the detection phase involved developing a lapse detection system based on EEG power spectral features such as power in the traditional bands (delta, theta, alpha, beta, etc.) and ratios between band powers.

In addition to power spectral features, several other signal processing techniques which were considered to have the potential to contribute towards reliable lapse detection were used to extract features from the EEG. These techniques were spectral coherence and asymmetry, fractal dimension, approximate entropy, and Lempel-Ziv complexity. Details of the signal processing algorithms and their lapse detection performance are presented in this chapter.

8.2 EEG feature extraction – an overview

Figure 8-1 shows a block diagram of the EEG feature extraction process. The process comprised the EEG acquisition stage, artifact removal stage, signal processing stage, and the feature matrix generation stage.

The EEG was recorded from electrodes at 16 scalp locations and digitized at 256 Hz (bandwidth 0.1–100 Hz) with a 16 bit A-D converter. The following standard bipolar derivations were used in the feature calculations: Fp1-F7, F7-T3, T3-T5, T5-O1, Fp2-F8, F8-T4, T4-T6, T6-O2, Fp1-F3, F3-C3, C3-P3, P3-O1, Fp2-F4, F4-C4, C4-P4, and P4-O2. Bipolar channels were preferred over referential channels as they reject common-mode noise better.

Next, each EEG channel was processed by rejecting epochs contaminated with artifacts, as preliminary work had shown that removal of artifacts from the EEG improved detector performance. Noise introduced by EEG artifacts may be counterproductive during model formation and classifier performance evaluation and, hence, reduce classifier effectiveness. As the first step of the artifact removal stage, the EEG was pre-processed using independent components analysis (ICA) to remove eye blink artifacts (Delorme and Makeig, 2004; Jung *et al.*, 2000a). The eye-blink artifact-free signal from each derivation was then filtered to remove 50 Hz mains activity using an IIR notch filter with a Q-factor of 35.

The mean and standard deviation of the first 2 min (baseline) of the signal were calculated. The signal was then transformed into z-scores relative to the baseline of the signal, thus enabling comparisons to be made between subjects and sessions. Two-s epochs containing samples with an absolute z-score >30 were rejected as artifacts and excluded from analysis in the signal processing algorithms described in the next section.

A *feature* is here defined as an arbitrary time series extracted from a single EEG derivation using a given signal processing algorithm. For example, if power spectral analysis is used to process the EEG, an extracted feature is the power in the alpha band, or ratio of power between delta and theta bands, over a set of consecutive epochs.

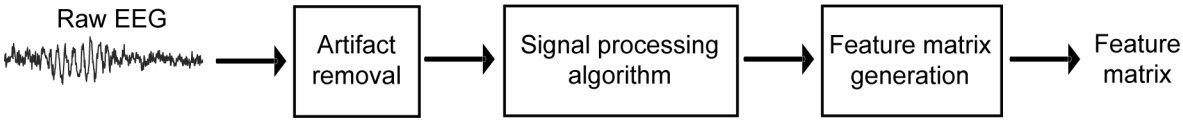

Figure 8-1. Block diagram of the EEG feature extraction process using various signal processing algorithms.

An epoch length of 2 s and an overlap of 1 s (50%) between successive epochs were used for all signal processing algorithms. The sliding process generated feature samples at a rate of 1 Hz, resulting in a 3600-element feature vector for a 1-h recording. The 2-s epoch length was chosen to obtain a reasonable degree of spectral resolution (where appropriate) and the overlap of 1-s was chosen to ensure reasonable temporal resolution (an estimate every second) for the features. This was important since a key requirement of the desired lapse detection system was its ability to detect short lapses (1~2 s).

8.3 Signal processing techniques for feature extraction

The power spectrum provides an estimate of power at each frequency of the signal. Several researchers have used EEG spectral power to detect changes in the level of alertness and arousal (Huang *et al.*, 2001; Jung *et al.*, 1997; Makeig and Inlow, 1993; Makeig and Jung, 1995; Makeig and Jung, 1996). Therefore, it seemed appropriate to begin the search for a reliable lapse detector by determining if there are EEG power spectral changes associated with lapses and, if so, determining the efficacy of a spectral-based lapse detection system.

Normal EEG is understood to have both linear and non-linear dynamic properties, leading to EEG patterns with different degrees of complexity (Natarajan *et al.*, 2004). Neuron voltage responses are non-linear – e.g., Hodgkin-Huxley model (1952) – and the EEG, which is generated by a combination of neurons firing together is likely to be non-linear. For example, it has been noted that "nonlinearity in the brain is introduced even at the cellular level, since the dynamical behaviour of individual neurons is governed by threshold and saturation phenomena" (Abasolo *et al.*, 2006). In fact, the human brain has been described as *the* most complex biological system and brain electrical activity has been found to exhibit complex non-linear behaviour (Koch and Laurent, 1999). Considering this fact, it was hypothesized that non-linear dynamical analysis techniques might prove a better approach to detect lapses than traditional linear methods (such as power spectral analysis) as they make better use of nonlinearities and dynamics in the EEG.

Progress made during recent years in non-linear dynamics theory has contributed new tools, useful in the analysis of the EEG (Elbert *et al.*, 1994). For example, non-linear analytical techniques have been used to investigate the EEG associated with various physiological and pathological states such as during meditation (Aftanas and Golocheikine, 2002), sleep and

slow-wave sleep (Ferri *et al.*, 1996; Kobayashi *et al.*, 2001), epilepsy (Elger *et al.*, 2000; Lehnertz, 1999), schizophrenia (Jeong *et al.*, 1998; Kim *et al.*, 2000), dementia and Parkinson's disease (Stam *et al.*, 1995), assessing the depth of sedation (Klonowski *et al.*, 2005), and Alzheimer's disease (Abasolo *et al.*, 2006). There is evidence to suggest that non-linear methods can be used to detect changes in the EEG that are not visible via visual observation or FFT (Le Van Quyen *et al.*, 2001).

Three non-linear methods were used to examine the EEG during lapses: fractal dimension (FD), approximate entropy (ApEn), and Lempel-Ziv complexity (LZ). The hypothesis was that the dynamics of the EEG change during lapses, and, that non-linear methods would be able to characterize such changes better than their linear counterparts. A summary of each algorithm, areas of application to date, and justification for their use to detect lapses is provided in §§ 8.3.4–8.3.6.

8.3.1 Power spectral analysis

Data in each 2-s epoch was first detrended to remove any linear trends (i.e., DC shifts) and the spectrum then estimated using a 40^{th}-order Burg model (Naidu, 1996). This parametric model method was selected to estimate power spectra due to its ability to provide a high degree of frequency resolution for short data records (Subasi, 2005a). It also ensures a stable autoregressive model and is computationally efficient. A high model order was found necessary to obtain adequate separation of the spectral bands of interest as lower order Burg models 'blurred' the spectrum, hindering the separation of spectral peaks in adjacent bands.

The spectral features listed in Table 8-1 were calculated. For a given epoch, the *spectral power* in each EEG band was calculated by finding the mean power across the band. Next, the *normalized power* was calculated for each band by dividing the spectral power in that band by the overall mean power across the entire spectrum. In addition, power ratios between bands were also calculated (Table 8-1). Power spectral analysis produced 13 spectral power (SP), 12 normalized spectral power (NSP), and 9 power ratio (PR) features per EEG derivation, giving a total of 34 features per derivation and 34×16 = 544 spectral features over the 16 derivations. A diagram illustrating how multiple features are produced from a single EEG derivation is shown in Figure 8-2.

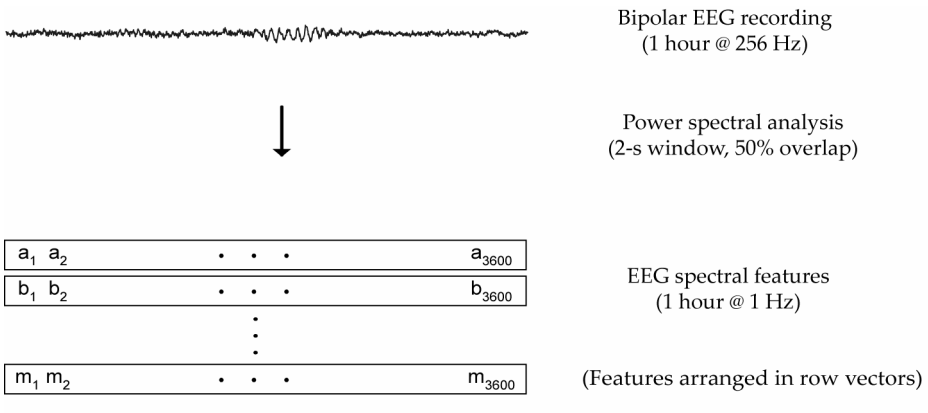

Figure 8-2 An example of the formation of multiple feature vectors from a single EEG derivation using power spectral analysis. Thirty four (13 spectral power, 12 normalized spectral power, 9 power ratio) feature vectors were extracted per derivation, with each vector being 1 by 3600.

Table 8-1 Spectral features calculated from each EEG derivation.

Feature	Frequency band
Mean spectral power	
Delta (δ)	1.0 – 4.5 Hz
Theta (θ)	4.5 – 8.0 Hz
Alpha 1 (α_1)	8.0 – 10.5 Hz
Alpha 2 (α_2)	10.5 – 12.5 Hz
Alpha (α)	8.0 – 12.5 Hz
Beta 1 (β_1)	12.5 – 15.0 Hz
Beta 2 (β_2)	15.0 – 25.0 Hz
Beta (β)	12.5 – 25.0 Hz
Gamma 1 (γ_1)	25.0 – 35.0 Hz
Gamma 2 (γ_2)	35.0 – 45.0 Hz
Gamma (γ)	25.0 – 45.0 Hz
High	> 45.0 Hz
All frequencies	Included only in SP
Spectral power ratios	
θ/β , θ/α, α/β, δ/θ, α/δ, β/δ, β_1/α, β_2/α, β_1/β_2	–

8.3.2 Spectral coherence

The spectral coherence estimate $C_{xy}(f)$ is a function of frequency, with values between 0 and 1, that indicates how well signal x linearly corresponds to signal y at each frequency. The coherence is a function of the power spectral densities (P_{xx} and P_{yy}) of x and y and the cross power spectral density (P_{xy}) of x and y, and is given by

$$C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)}. \quad (8\text{-}1)$$

Spectral coherence was calculated using the MATLAB routine `mscohere()`. $C_{xy}(f)$ is also known as the *magnitude-squared coherence* function. The routine provided the magnitude-squared coherence estimate $C_{xy}(f)$ of two input signals x and y using Welch's averaged, modified periodogram method. Coherence was calculated for all 120 pair-wise combinations of the 16 EEG channels with an estimate of $C_{xy}(f)$ for each of the 12 frequency bands (see Table 8-1) to give a total of 1440 coherence features.

It has been shown that coherence increases within and between sensory and motor sites during task performance (Aoki *et al.*, 1999). Therefore, the hypothesis for this work is that lapses in task performance are reflected in decreases in coherence across specific brain sites and that this information could be incorporated in the detection of lapses.

8.3.3 Spectral amplitude asymmetry

Another hypothesis was that EEG activity in different cortical regions desynchronizes during lapses. This in turn may be reflected in terms of synchrony or changes in spectral asymmetry between different EEG sites. Synchrony has been shown to be of no value for lapse detection by another member of the Lapse Research Programme (Begg, 2003). Therefore, spectral asymmetry was investigated as a candidate that could potentially provide additional information regarding lapses to complement spectral power measures.

Spectral asymmetry is defined as the ratio of differences in power in a given EEG band between two scalp locations (A and B), and was calculated using $(A–B)/(A+B)$, where A and B are the spectral band powers recorded from two different bipolar derivations; amplitude asymmetry is zero when $A = B$. Comparisons were made for left *vs*. right electrode locations (8 pairs): Fp1 *vs*. Fp2, F7 *vs*. F8, F3 *vs*. F4, T3 *vs*. T4, C3 *vs*. C4, T5 *vs*. T6, P3 *vs*. P4, and O1 *vs*. O2 and anterior *vs*. posterior (6 pairs): F8 *vs*. T6, Fp2 *vs*. O2, F4 *vs*. P4, Fp1 *vs*. O1, F3 *vs*. P3, and F7 *vs*. T5, a total of 14. Each comparison assessed the asymmetry in one of the 12 bands listed in Table 8-1 as well as for all bands combined. This resulted in $14 \times 13 = 182$ asymmetry features.

8.3.4 Fractal dimension

A fractal is a shape which retains similarity in structural detail despite magnification (scaling). The complexity of the structure of such a set, invariant under this scaling, can be characterized by a single number: the fractal dimension (FD). Correlation dimension

(Grassberger and Procaccia, 1983b) has also been used to measure changes in the EEG during sleep (Achermann *et al.*, 1994; Kobayashi *et al.*, 2001) and distinguishing Alzheimer patients from healthy individuals via the EEG (Jelles *et al.*, 1999). However, calculation of the correlation dimension requires the reconstruction of a 'strange attractor' in a multi-dimensional phase space (Accardo *et al.*, 1997). FD has the advantage that it allows the measurement of the complexity of a signal to be evaluated directly on the time domain without reconstruction and, hence, provides a direct link between EEG variations and complexity changes (Accardo *et al.*, 1997).

The brain has been interpreted as a non-linear dynamical system whose state can be described by self-similar curves (Lutzenberger *et al.*, 1995). EEG signals are an example of such curves and their complexity, as estimated by FD, has been shown to correspond to different physiopathological conditions (Accardo *et al.*, 1997). The FD of any signal varies between 1 and 2: the more complex a waveform, the higher its FD. FD has been shown to be effective as a means of comparing differences in the complexity of EEG signals recorded from patients with bipolar mood disorder and controls (Bahrami *et al.*, 2005) and in the analysis of epileptic ictal events (Bullmore *et al.*, 1994).

Higuchi's algorithm (Higuchi, 1988) was used to estimate the fractal dimension (FD) of each EEG derivation because it is computationally efficient and also provides a stable estimate of FD using a lower number of samples of data ($N \geq 125$) compared to other FD algorithm implementations (Accardo *et al.*, 1997). This allowed the FD to be estimated with the same temporal resolution as other features.

The Higuchi algorithm generates multiple time series from a given EEG segment consisting of N samples, $x(i)$, $i = 1,\ldots,N$. Starting from the m^{th} data point of the signal, points are sampled with scale size k such that successive samples are k data points apart. This process is repeated for different starting points ($m = 1,...,k$) and different scale sizes ($1,...,k_{\max}$) such that, for each scale size k, a set of time series is obtained,

$$x_m^k = \{x(m), x(m+k), \ldots, x(m+\left\lfloor \frac{N-m}{k} \right\rfloor k)\}, \tag{8-2}$$

where $\lfloor a \rfloor$ denotes the integer part of a.

For each time series x_m^k, the absolute differences between each two successive data points are summed to calculate the vertical length of the signal measured with the scale size k, starting at the m^{th} data point:

$$L_m(k) = \frac{\sum_{i=1}^{\lfloor (N-m)/k \rfloor} \left| x(m+ik) - x(m+(i-1)k) \right| (n-1)}{\left\lfloor \frac{(N-m)}{k} \right\rfloor k}. \quad (8\text{-}3)$$

The length of the EEG segment $L(k)$ is the mean of values, for $m = 1, ..., k$. Note that $(n-1)/\left(\left\lfloor \frac{(N-m)}{k} \right\rfloor k\right)$ is a normalization term (Accardo *et al.*, 1997). This procedure is repeated for each k ranging from 1 to k_{max} yielding a sum of average lengths $L(k)$ for each k:

$$L(k) = \sum_{m=1}^{k} L_m(k) \quad (8\text{-}4)$$

If the value of $L(k)$ is proportional to k^{-d}, the signal is considered fractal-like with a FD of d. The angular coefficient of the linear regression (i.e., slope) of the graph $\log(L(k))$ *vs*. $\log(1/k)$ gives an estimate of d.

Parameters suggested by Accardo *et al*. (1997) were used for estimating the FD of the EEG ($k_{max} = 6$). The FD was estimated for each EEG derivation, resulting in 16 FD feature vectors per session with a value of FD every 1.0 s.

8.3.5 Approximate entropy

Entropy is a concept that addresses system randomness and predictability (Grassberger and Procaccia, 1983a). It quantifies the predictability of amplitude values of a signal, based on the knowledge of amplitudes of previous samples (Bruhn *et al.*, 2000).

The approximate entropy (ApEn) measure was developed to quantify the amount of regularity in a signal without *a priori* knowledge about the generating system (Pincus and Goldberger, 1994), as cited in (Zhang and Roy, 2001). ApEn is non-negative, with a larger number indicating more irregularity, unpredictability, and randomness of the raw signal (Zhang and Roy, 2001). It is also relatively unaffected by low level noise, is robust to occasional very large or small artifacts, and gives meaningful information with a reasonable number of data points, properties which make it appealing for characterizing changes in the complexity of the EEG over time.

For a perfectly regular data series, knowledge of prior values enables the subsequent value to be predicted correctly. This has an associated ApEn measure of 0. However, with increasing irregularity, even with knowledge of the previous values, the prediction of the subsequent value will be worse, leading to an increased ApEn value. Thus, the subsequent

value in an entirely irregular data series cannot be predicted accurately from knowledge of previous data values.

ApEn has been used in a variety of contexts including identification of fetal distress from heart rate (Ferrario *et al.*, 2006), prediction of paroxysmal atrial fibrillation (Shin *et al.*, 2006), human respiratory variability (Burioka *et al.*, 2003), estimation of depth of anesthesia (Zhang and Roy, 2001) and anesthetic drug effect (Bruhn *et al.*, 2000), and differentiating between sleep stages (Burioka *et al.*, 2005).

Two parameters are used in the ApEn algorithm. They are (a) the embedding dimension m and (b) the tolerance of the noise filter r. The embedding dimension m specifies the number of previous values used for the prediction of the subsequent value and the noise filter value r is expressed as a proportion of the standard deviation of the amplitude values of the n samples in the data sequence (Bruhn *et al.*, 2000). Generally, a value of $m = 2$ and r in the (0.1~0.25) SD range is suggested in the literature (Pincus *et al.*, 1991).

Assuming that the raw EEG samples are $x_1, x_2, \cdots, x_n$, and n is the total number of samples in the sequence, ApEn of an EEG sequence is calculated as follows (Bruhn *et al.*, 2000; Zhang and Roy, 2001):

1. Construct m-vectors $X(1)$ to $X(n-m+1)$ defined by $X(i) = [x(i), x(i+1), ..., x(i+m-1)]$, for $i = 1, 2, ...n-m+1$.
2. Define $d[X(i), X(j)]$, the distance between $X(i)$ and $X(j)$, as the maximum absolute difference between their corresponding scalar elements, i.e.,

$$d[X(i), X(j)] = \max_{k=0}^{m-1} \left[\left| x(i+k) - x(j+k) \right| \right]. \tag{8-5}$$

3. For a given $X(i)$, count the number of j ($j = 1, 2, ..., n-m+1, j \neq i$) such that $d[X(i), X(j)] \leq r$, denoted as $n^m(i)$. Then, for $i = 1, 2, ..., n-m+1$, define $C_r^m(i) = n^m(i)/(n-m+1)$.
4. Take the natural logarithm of each $C_r^m(i)$, and average it over i, i.e.,

$$\phi^m(r) = \frac{1}{n-m+1} \sum_{i=1}^{n-m+1} \ln C_r^m(i). \tag{8-6}$$

5. Increase dimension to $m+1$. Repeat steps (1) to (4) and find $C_r^{m+1}(i)$ and $\phi^{m+1}(r)$.

Theoretically, ApEn is defined as

$$ApEn(m,r) = \lim_{n\to\infty}\left[\phi^m(r) - \phi^{m+1}(r)\right]. \quad (8\text{-}7)$$

However, in practice, the number of data points (n) is finite, and the result obtained through the above steps is the estimate of ApEn when the data length is n, which can be denoted as

$$ApEn(m,r,n) = \phi^m(r) - \phi^{m+1}(r). \quad (8\text{-}8)$$

Values suggested in the literature for m and r (2 and 0.2, respectively) were used in the ApEn algorithm (Bruhn *et al.*, 2000; Pincus, 1995; Pincus *et al.*, 1991; Zhang and Roy, 2001). ApEn was estimated for each EEG derivation, resulting in 16 ApEn feature vectors per session with an ApEn value calculated every 1.0 s.

8.3.6 Lempel-Ziv complexity

Lempel-Ziv (LZ) complexity (Lempel and Ziv, 1976) provides a non-parametric measure of complexity of a one-dimensional signal, such as the EEG. Its advantages are that it is simple to compute, does not require long data segments to be effective, and is more effective for real-time EEG processing (Radhakrishnan and Gangadhar, 1998; Zhang and Roy, 1999; Zhang *et al.*, 1999) compared to other complexity measures such as correlation dimension (Yaylali *et al.*, 1996) and neural complexity (Tononi *et al.*, 1994). It has been useful in quantifying the depth of anesthesia (Zhang and Roy, 2001), predicting epileptic seizures (Radhakrishnan and Gangadhar, 1998), and analysing the dynamical behaviour of the background EEG of patients with Alzheimer's disease (Abasolo *et al.*, 2006).

Lempel and Ziv proposed that the complexity of a finite sequence could be evaluated from the point of view of a "simple self-delimiting learning machine which, as it scans a given digit sequence $S = s_1 s_2 \cdots s_n$ from left to right, added a new word to its memory every time it discovered a sub-string of consecutive digits not previously encountered". The complexity counter $c(n)$ is increased by one unit each time a new sub-string of characters is encountered along S (Lempel and Ziv, 1976; Radhakrishnan and Gangadhar, 1998). Only two operations are permitted in the construction of a string: copying old patterns and inserting new ones (Zhang and Roy, 2001).

Prior to calculating $c(n)$, the raw EEG signal $(x_1, x_2, \cdots x_n)$ must be transformed into a finite symbol string $S = s_1 s_2 \cdots s_n$, where $s_i \in \{0,1\}$ and n is the length of the EEG segment (i.e., window length). Firstly, the mean of the EEG signal $x_m = (1/n)\sum_{i=1}^{n} x_i$ was calculated. Then each x_i was compared with x_m and transformed into a binary string $S = s_1 s_2 \cdots s_n$: $s_i = 0$ if $x_i < x_m$, or $s_i = 1$ otherwise. Then $c(n)$ can be estimated from the string S as follows (Abasolo *et al.*, 2006; Zhang and Roy, 2001; Zhang *et al.*, 1999):

1. Let Q and R denote two sub-strings of S, and QR be the concatenation of Q and R. In addition, let $QR\pi$ be the result of deleting the last character from QR, where π denotes the operation to delete the last character in the sequence. Let $v(QR\pi)$ denote the vocabulary of all different sub-strings of $QR\pi$. At the start, $c(n)$=1, $Q = s_1$, $R = s_2$, therefore $QR\pi = s_1$.
2. In general, $Q = s_1 s_2 \cdots s_r$, $R = s_{r+1}$, then $QR\pi = s_1 s_2 \cdots s_r$; if R belongs to $v(QR\pi)$, then R is a sub-string of $QR\pi$ and, hence, not a new sequence.
3. Renew R to be $s_{r+1}s_{r+2}$ and assess whether it belongs to $v(QR\pi)$.
4. Repeat steps (2)-(3) until R does not belong to $v(QR\pi)$. Now $R = s_{r+1}s_{r+2} \cdots s_{r+i}$ is not a sub-string of $QR\pi = s_1 s_2 \cdots s_{r+i-1}$, so increase $c(n)$ by 1.
5. Thereafter, combine Q with R, and Q is renewed to be $Q = s_1 s_2 \cdots s_{r+i}$, and $R = s_{r+i+1}$.
6. Repeat above procedure until R is the last character. At this point, the number of different sub-strings of S is $c(n)$, its complexity measure.

In order to obtain a complexity measure independent of string length n, $c(n)$ must be normalized to $C(n)$. It has been shown that this can be achieved via $b(n)$, where $b(n) = n/\log_\alpha(n)$ and α denotes the number of different symbols in the string (Shaw *et al.*, 1999). For a binary sequence, $\alpha = 2$. The normalized complexity measure is given by

$$C(n) = \frac{c(n)}{b(n)} \tag{8-9}$$

and reflects the recurrence rate of new patterns along the sequence and thus captures the temporal structure of the signal (Zhang and Roy, 2001).

LZ complexity features were estimated for each EEG derivation, resulting in 16 LZ complexity feature vectors per session with a value calculated every 1.0 s.

8.4 EEG feature matrix

8.4.1 Assembling the feature matrix

A *feature matrix* for a session's EEG data was created by grouping various combinations of EEG features calculated using the signal processing algorithms described in § 8.3. This was achieved by placing m feature vectors (each of length n) as row vectors in the feature matrix as shown in Figure 8-3.

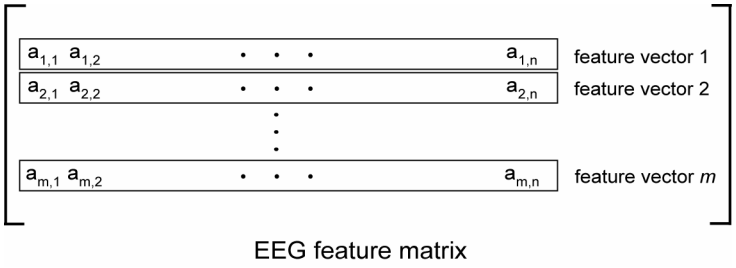

Figure 8-3 Diagram depicting the formation of the EEG feature matrix by placing feature vectors of length-*n* to form an *m* by *n* feature matrix. For a 1 hour session, there were *n*=3600 samples, and *m* varied depending on which features were selected to create the classification model.

For algorithms which produced 1 feature per EEG derivation (such as FD, ApEn and LZ), the generated EEG feature matrix was of size 16 by n. For example, the EEG feature matrix based on FD measures had 16 FD feature vectors as row vectors in the matrix. However, for measures such as power spectra, which had multiple features per EEG derivation, the size of the EEG feature matrix was much larger. These were arranged in the EEG feature matrix as rows, in order of EEG derivation. That is, all the spectral features for the first derivation were listed in the first 34 rows of the feature matrix, followed by the features of the second derivation from rows 35-69, etc., until all features from all derivations were entered into the feature matrix. This resulted in 544 (34×16) feature vectors in the matrix for power spectra.

8.4.2 Principal component analysis of the feature matrix

Since data from 16 EEG derivations were used to form the feature matrix, it was highly likely that features in adjacent EEG derivations would be highly correlated with each other. The feature matrix would therefore contain redundant information. Principal component analysis (PCA) was used to transform the feature vectors into orthogonal components (and, hence, reduce the redundancy within the original features) to aid the formation of the classification models.

PCA has been used in a wide range of research areas as a non-parametric method of extracting relevant information from complex and often confusing datasets (Jolliffe, 2002). It is often used to reduce high-dimensional complex data into a much simpler and lower dimensional form, which can reveal structures underlying a system that may otherwise be hidden. PCA achieves this whilst retaining all of the characteristics contributing to the variance in the original data.

In PCA, the original feature matrix Y (m features, n samples) is transformed into a set of principal components (PCs) X (also $m \times n$), which are related to Y by a linear transformation P such that $PY = X$. Geometrically speaking, P is a rotation and stretch which transforms

Y into X. Conversely, the original feature matrix Y can be reconstructed from the PCs by pre-multiplying X by the inverse transformation matrix P^{-1}.

The PCs are ranked in order of descending importance in terms of the amount of variance explained. Therefore, it is possible to reduce the dimensionality of the data by using the first p out of m PCs ($p < m$) without significant loss of information, provided p is chosen appropriately.

8.5 Lapse index

Following the extraction of EEG features, creation of the feature matrix, and transforming feature data using PCA, the next step was to determine if changes in the feature vectors correlated with an independent measure of whether or not the subjects were lapsing, i.e., a *lapse index* (LI).

Chapter 7 presented four candidates that could be used for the LI, such as flat spots, video BMs, lapses of responsiveness (simply referred to as 'lapses' in this chapter), and definite BMs. From a detection perspective, it is pertinent to identify brief episodes during which a subject unintentionally stops responding to a task, since the occurrence of such events during real-life tasks such as driving has the potential to cause accidents. As stated in Chapter 7, video BMs and tracking flat spots were both independent and conservative measures of lapses of responsiveness.

However, results in Chapter 7 also showed that tracking flat spots and video BMs exhibited a relatively low mean correlation of $r = 0.40 \pm 0.06$ (mean ± SE) over the 8 subjects. Video sleep events were more abundant than flat spots due to the latter being an overly conservative estimate of lapses occurring during tracking. For example, there were numerous instances in the data where the subject was briefly asleep according to the video, and their tracking of the target was incoherent but these were not counted as flat spots and, hence, flat spots represented a subset of tracking lapses. Note that it was decided not to include incoherent tracking instances as lapses due to their being less definitive and the difficulty in defining their start and end points – i.e., these incoherent tracking instances had less certainty and poorer temporal resolution compared to flat spots.

It was also noted in Chapter 7 that the CTT data contained a sub-category of lapses in which clear behavioural signs of sleep (such as eye-lid closure and head-nodding) and tracking flat spots overlapped. These events were termed *definite behavioural microsleeps* (BMs) to emphasize their arousal-related nature and distinguish them from EEG-defined microsleeps (Ogilvie, 2001). The term 'definite' was used to refer to these events, emphasizing that they

were identified using two independent and conservative measures. Note that a video BM may or may not have had a simultaneous flat spot in tracking and, therefore, could not, on its own, be called 'definite'. It was also emphasized that definite BMs were a highly conservative measure and missed many genuine lapses.

Ideally, a lapse detection device should detect both lapses in attention and lapses caused by low arousal, because, despite presumably different underlying mechanisms in the brain, both types of lapse cause a subject to be unresponsive for brief periods of time, thereby potentially placing themselves and others in danger. Therefore, lapses, as defined by the presence of either a video BM and/or a flat spot, were selected as the events to be detected by an EEG-feature-based lapse detector. Detector performance was also assessed for systems trained to detect flat spots, video BMs, and definite BMs. However, results showed that a system based on detecting lapses performed best and, hence, only results based on detecting lapses are reported here. The LI was generated at a frequency of 1 Hz.

8.6 Classification models to detect lapses from EEG features

The next step in the design of the lapse detection system was to train a *classification model* capable of detecting lapses in new subjects, using data from their feature matrices. The process involved forming a classification model, based upon linear discriminant analysis (LDA), using PCs extracted from the feature matrix as predictive variables and LI as the grouping variable. In this section, a method of using data from a single subject and session is presented. The combination of data from more than one session and subject is explained in later sections.

8.6.1 Linear discriminant analysis

LDA is a statistical technique used to determine which continuous variables can discriminate between two or more groups (Ripley, 1996) and was used to form classification models capable of detecting lapses. The classical method of linear discrimination for two groups was first described by Fisher (1936) and was achieved by maximizing the ratio of the *between-group variance* to the *within-group variance*.

LDA assumes that the group memberships of the initial cases (training set) are known correctly. This analysis yields information which can then be used to classify a future case with an unknown group membership, into a group.

A brief overview of LDA is as follows. Refer to (McLachlan, 2005; Ripley, 1996) for a more complete coverage. Assume that X (the matrix of PCs derived from the EEG feature matrix

Y) denotes a $m \times n$ matrix of samples where *m* is the number of variables (or features), *n* is the number of samples, and *g* denotes the number of groups (or classes). In the lapse detection framework, samples fall into either the 'lapse' or 'not lapse' groups (*g*=2). Let *W* denote the within-class covariance matrix (i.e., covariance matrix of the variables centered on the group mean) and *B* denote the between-groups covariance matrix (i.e., predictions by the group means). Furthermore, let *M* denote the $g \times n$ matrix of group means, and *G* denote the $m \times g$ matrix of group indicator variables (such that $g_{ij} = 1$ if case *i* is assigned to group *j* and $g_{ij} = 0$ otherwise) such that the predictions are given by the multiplication of *G* and *M*, GM. Finally, let $\overline{X}$ be a $m \times n$ matrix whose columns are all given by the row means of *X*. The sample covariance matrices are

$$W = \frac{(X - GM)^T (X - GM)}{n - g} \text{ and} \tag{8-10}$$

$$B = \frac{(GM - \overline{X})^T (GM - \overline{X})}{g - 1}, \tag{8-11}$$

where *B* has rank at most $\min(n, g-1)$.

Fisher introduced a technique to find a $1 \times m$ vector $\mathbf{a}$ of scalar coefficients, which is used to linearly combine the variables (i.e., $\mathbf{a}X$) to maximize the ratio of the *between-group variance* to the *within-group variance* of *X*. Fisher showed that this is equivalent to maximizing the ratio $\mathbf{a}^T B\mathbf{a} / \mathbf{a}^T W\mathbf{a}$. Solving for $\mathbf{a}$ yields the *discriminant coefficients* (or *canonical variates)* of the classification function, which can be used to classify new cases (Fisher, 1936; Ripley, 1996).

Fisher's linear discriminant function for the two group scenario (e.g. lapse *vs*. no lapse) is given by

$$Z = a_1 x_1 + a_2 x_2 + ... + a_m x_m, \tag{8-12}$$

where $x_1, x_2, ..., x_m$ are the feature row vectors (predictive variables) and $a_1, a_2, ..., a_m$ are the discriminant coefficients which form $\mathbf{a}$. The optimal dividing point *C* between the two groups is given by

$$C = \frac{\overline{Z}_1 + \overline{Z}_2}{2} + \ln\left(\frac{q_2}{q_1}\right) \tag{8-13}$$

where the two group mean values of Z are denoted by $\overline{Z}_1$ and $\overline{Z}_2$, and the prior probabilities of belonging to group 1 or 2 are denoted by q_1 and q_2, respectively. Note that q_1 and q_2 are

estimated by calculating the proportion of lapses and non-lapses from the LI. A new case is assigned to group 1 if $Z < C$ and to group 2 otherwise.

The raw values of $a_1, a_2, ..., a_m$ are not directly comparable to enable the determination of the contribution of each PC towards classification. An estimate of the relative contribution of each variable on the discriminant function is obtainable from the *standardized discriminant coefficients b*. These coefficients are calculated using the formula

$$b_i = a_i \sqrt{W_{ii}} \tag{8-14}$$

where *W* is the within-groups covariance matrix as defined in Eq. 8-10 (Afifi and Clarke, 1998; Mueller and Cozad, 1993).

8.6.2 Forming the classification model

The lapse indices and EEG feature matrices from both sessions of each subject were concatenated to form a single feature matrix ($m \times 7200$) and LI (1×7200) per subject. These data were then used to form a classification model for each of the 8 subjects, resulting in 8 models. The process used to create each model is illustrated in Figure 8-4 and summarized as follows.

Firstly, the mean over the entire length of the record was calculated for each vector of the feature matrix. The means were then subtracted from the feature vectors to produce zero-mean vectors in the feature matrix. This was a necessary prerequisite for PCA. Following this, PCA was performed on the mean-subtracted $m \times n$ feature matrix to derive *m* PCs, each of length *n*. Next, the PCs were converted to z-scores by subtracting the overall means and dividing by the standard deviations. The z-score transformed PCs and the LI were used to form a linear discriminant classification model for each subject.

The discriminant analysis toolbox for MATLAB®, written by Michael Kiefte (1999), was used for the analysis. As seen in Figure 8-4, the classification model for each subject consisted of (a) linear discriminant function coefficients and (b) a transformation matrix generated via PCA.

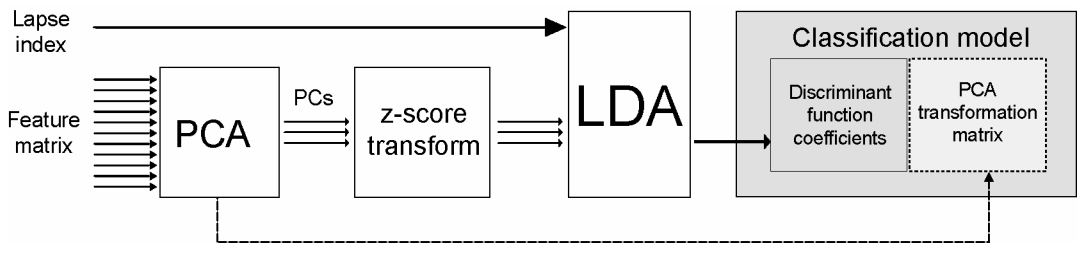

Figure 8-4 A block diagram depicting the creation of a lapse classification model.

8.7 Combining multiple classification models to form an overall detection model

It has been shown that, in general, combining the output of several models tends to increase predictive performance over a single model (Witten and Frank, 2000). Therefore, rather than using the output of a single overall classification model, a superior method is to combine the output of all classification models and arrive at a consensus. Strategies previously used to combine multiple models include the techniques of *bagging*, *boosting*, and *stacking* (Witten and Frank, 2000).

8.7.1 Bagging

The bagging method (Brieman, 1996a) resamples datasets from the original data and uses these training sets to generate the models. However, since these datasets are generated by resampling the same original data, they are not independent of each other. During classification of a new test dataset, each model gets a "vote" regarding the group of a new test dataset and the group which receives the most votes from all individual models is selected as the correct one. For situations where numerical prediction is required, the mean of all models is taken as the overall prediction. Bagging has been found to produce a combined model that often performs substantially better than a single model built from the original data, and is never substantially worse (Witten and Frank, 2000). In bagging, each model receives equal weight.

8.7.2 Boosting

In contrast to bagging, boosting (Schapire *et al.*, 1997) alters the weights of the models to give more influence to the more successful ones. It tries to seek models that complement one another. Like bagging, boosting also uses voting (for classification) and averaging (for prediction) to combine the output of individual models. Both bagging and boosting are usually used to combine models of the same type. However, in bagging, the individual models are built separately, whereas in boosting, each new model is influenced by the performance of those built previously.

8.7.3 Stacked generalization

Stacked generalization (or simply 'stacking') (Wolpert, 1992) is an alternative approach to combining multiple models and was chosen to combine the outputs of the multiple lapse classifiers in this study. This was done to overcome a substantial problem with the voting procedure in bagging and boosting which does not clarify which base models to trust.

Stacking, however, aims to determine how to best combine the base models via an additional *meta-learner* algorithm (Figure 8-5).

Despite its advantages, stacking is less widely used than bagging or boosting, partly because it is difficult to analyse theoretically and because there is no generally accepted optimal methodology for performing it. A study by Ting and Witten (1999) found that the performance of bagging, boosting, and stacking were "very competitive". They found that stacking performed better than bagging or boosting on 3 out of 6 datasets but performed the worst on two of the datasets (small real-world dataset). They state that this is not surprising since cross-validation inevitably produces poor estimates for small datasets. Another downside of stacking is the increased computation time required to process all level-0 classifiers (instead of a single classifier) to produce a classification (Kotsiantis and Pintelas, 2004). It is also difficult for one to perceive the underlying reasoning process leading to a decision due to the multiple level-0 models, thus reducing comprehensibility (Kotsiantis and Pintelas, 2004).

The outputs of the base models (also known as *level-0 models*) were fed as the inputs to the meta-learner (*level-1 model*). During the classification phase of the stacked learner, new cases were fed into the level-0 models, each producing a classification value at their output. These level-0 predictions were then fed into the level-1 model which combined them linearly by scaling the output of each model by its weight, summing the scaled model outputs, and applying a threshold to the summed output to obtain an overall prediction (Figure 8-6).

It has been suggested that some of the test data be held back and used to train the level-1 model, with the level-0 models being trained on the remaining data (Witten and Frank, 2000). Once the level-0 models are trained, the holdout data are classified using the level-0 models, which then form the training data for the level-1 model. Since the holdout data were not used to train the level-0 models, their predictions are unbiased and, therefore, the level-1 training data accurately reflect the true performance of the level-0 models. However, the downside of the holdout method is that it deprives the level-1 model of some of the training data. This problem was overcome by applying 8-fold cross-validation which ensured that all of the training data were used to train the level-1 model. Each instance of the training data was used in one test-fold of the cross-validation and the predictions from the models built from the corresponding training fold were used to build the level-1 training set. This generated a level-1 training set for each level-0 training set.

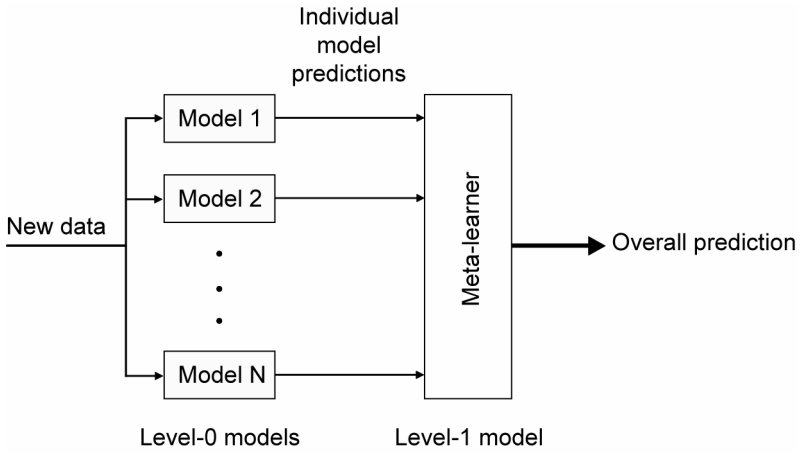

Figure 8-5 Schematic of the stacked generalization approach combining the output of several classifiers for overall prediction.

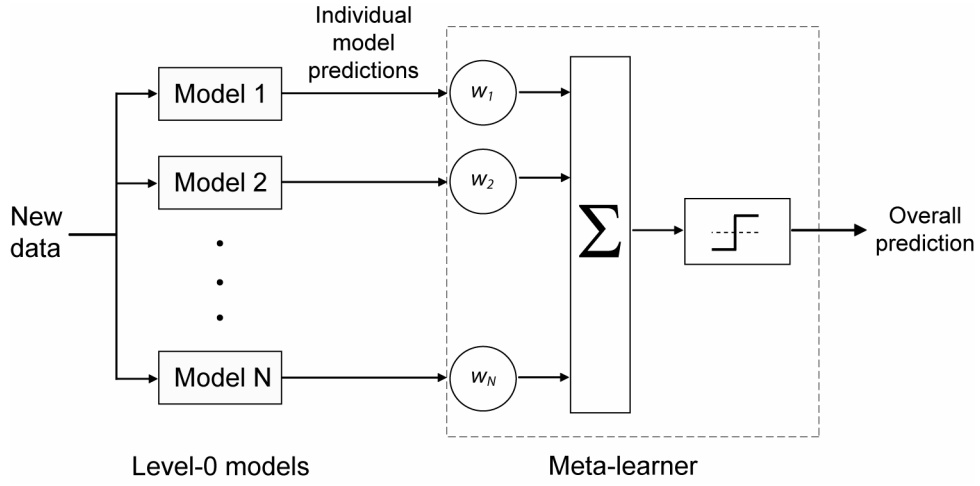

Figure 8-6 Diagram showing internal structure of the meta-learner used in the stacked generalization approach. The model weights are depicted by w_1, w_2, …, w_N and these are used by the meta-learner to scale the level-0 model outputs before combining them to form an overall prediction.

8.8 Overall detection model validation

The following steps were followed to validate the overall lapse detection model:

1. Reserve one of the 8 subjects as the *validation subject* and put their data aside.
2. Create classification models using data from the 7 remaining subjects.
3. Select one subject from the 7 subjects (*test subject*) in '2' and feed their features into the 6 level-0 models (excluding their own) for classification. This yields 6 level-0 outputs which are stored in a matrix. Note that the 'raw' output of the classifiers (i.e.,

continuous values between 0.0 and 1.0 indicating the probability of a given sample being a lapse) was used in the steps that follow.

4. Determine the meta-learner weights for the 6 level-0 model outputs by linearly combining them to estimate the LI of the test subject. Constrained least-squares fitting (coefficients restricted to >0) was used to combine the output of the 6 models. This approach minimized the least-squares error between the combined output (i.e., meta-learner output) and LI. It produced a set of positive regression coefficients for the 6 models, with larger coefficients associated with models contributing a greater degree towards the meta-learner output. These coefficients were stored in a matrix.
5. Determine the optimal threshold value required to be applied to the meta-learner output to obtain a binary classification (i.e., lapse/not lapse) by selecting the threshold that yields the maximum phi correlation between the meta-learner output and LI.
6. Repeat steps '3' to '5' until all 7 subjects are used as test subjects.
7. Calculate mean *meta-learner weights* and mean *meta-learner output threshold* by averaging over the 7 test subjects.
8. Feed the validation subject's data to all 7 level-0 models in the stacked generalization system and obtain the final prediction from the meta-learner output. The meta-learner scales the individual predictions of the level-0 models by the weights calculated in '7', sums predictions of all level-0 models, and finally applies the output threshold also determined in '7' to provide a final (binary) prediction of lapse (1) or no-lapse (0).
9. Calculate the correlation between the validation subject's LI and the meta-learner output after applying the mean output threshold. The correlation measure used was the phi correlation coefficient (ϕ) (Sheskin, 1997) with each validation subject's phi coefficient denoted by ϕ_v.
10. Repeat steps '1' to '9' and obtain ϕ_v for each of the 8 subjects.
11. Calculate the mean across all 8 values of ϕ_v to give the overall detector performance.

Performance of the lapse detector was evaluated using several metrics. The primary performance metric was the mean phi correlation, as described above. In addition, two other performance measures (which are independent of operating point) were also calculated: (a) area under the receiver-operator characteristic curve (AUC-ROC); and (b) area under the precision-recall curve (AUC-PR). These calculations were performed using the ROCR package (Sing *et al.*, 2005). The performance of the classification models is quoted using all three measures.

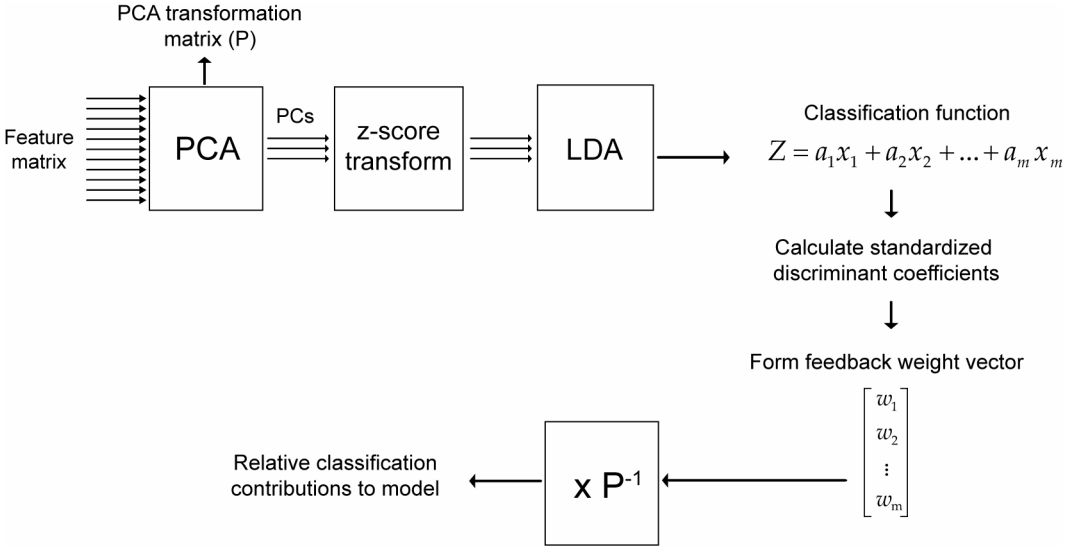

Figure 8-7 Diagram of the process used to determine the relative contributions towards discrimination by all features used in the classification model.

The effect of using uniform meta-learner weights on the performance of the overall lapse detector was also investigated. Note that this process is equivalent to removing the stacked generalization section from the system described above.

8.9 Contribution of features to discrimination

After determining the performance of the overall detection model via cross-validation, the next step was to determine the amount each feature contributed to the overall discrimination ability of the model. Figure 8-7 shows the procedure used. The aim of this analysis was to determine if any redundant features could be excluded.

Firstly, the discriminant coefficients of each classification model were converted to standardized discriminant coefficients as described in § 8.6.1. These coefficients were then normalized and used to form a *feedback weight vector*. This contained m elements, corresponding to the PC features used to construct the classifier. That is, the feedback weight vector's elements indicated the relative contribution towards lapse classification of each of the PCs. However, as each PC was a linear combination of all input features, it was possible to translate the feedback weight vector back to feature space to determine the relative contribution of each of the original features to the classification model. This was achieved by multiplying the feedback weight vector by the inverse of the PCA transformation matrix (P^{-1}) calculated during PCA at the model formation phase. This procedure provided the *relative contribution* of each feature towards the classification power of a particular level-0 model. The procedure was repeated to calculate the relative contribution of features in all level-0 models. However, as mentioned earlier (see § 8.7.3), the

generalization performances of the level-0 models were not equal and hence the relative contributions of each level-0 model were adjusted according to the model weights. These scaled contributions were then summed to obtain the contribution of features towards the overall detection model.

8.10 Results

The performance of the lapse detector can be measured both in terms of its ability to detect the *lapse state* (in 1-s epochs) and *lapse events*. The majority of this chapter is devoted to evaluating its ability to detect the former. This section contains results for the performance of lapse detectors based on the spectral measures introduced earlier in the chapter, the three non-linear measures, and a combination of spectral and non-linear feature-based lapse detectors. Detector performance is presented using several metrics: mean phi correlation, area under the receiver-operator characteristic curve (AUC-ROC) and, area under the precision-recall curve (AUC-PR). The performance of each classification model is quoted using all three measures. Results on the amount of contribution of EEG features and derivations to the best overall lapse detection model are presented. The variation in performance with the number of PCs used to form the detector model is also presented.

8.10.1 Detector performance – spectral measures only

Table 8-2 provides a summary of system performance for a lapse detector based on spectral power (SP), normalized spectral power (NSP), power ratios (PR), and combinations thereof. Detector performances are shown for the two weighting schemes (uniform and least-squares) used to scale the classifier outputs to obtain the overall prediction. Detectors based solely on SP, or incorporating SP with NSP, and utilizing least-squares weights for combining the classifier outputs (to obtain the overall prediction) provided the best generalization performance (ϕ~0.39). This fact is confirmed by observing that SP and SP+NSP based detectors had the largest AUC-ROC and AUC-PR values as shown in Table 8-3. As mentioned earlier, AUC-ROC and AUC-PR being threshold-independent measures, emphasizes that SP and SP+NSP indeed give the best performance out of the 7 spectral-feature-based detectors. Figure 8-8 shows the mean ROC and PR curve for the spectral power (SP) based detector.

Figure 8-9 shows the mean meta-learner weights (normalized) of the spectral-power-based lapse detector. The mean weights were calculated by averaging the meta-learner weights over all validation runs. Results show that 4 out of 8 models were assigned a normalized mean weight over 0.1.

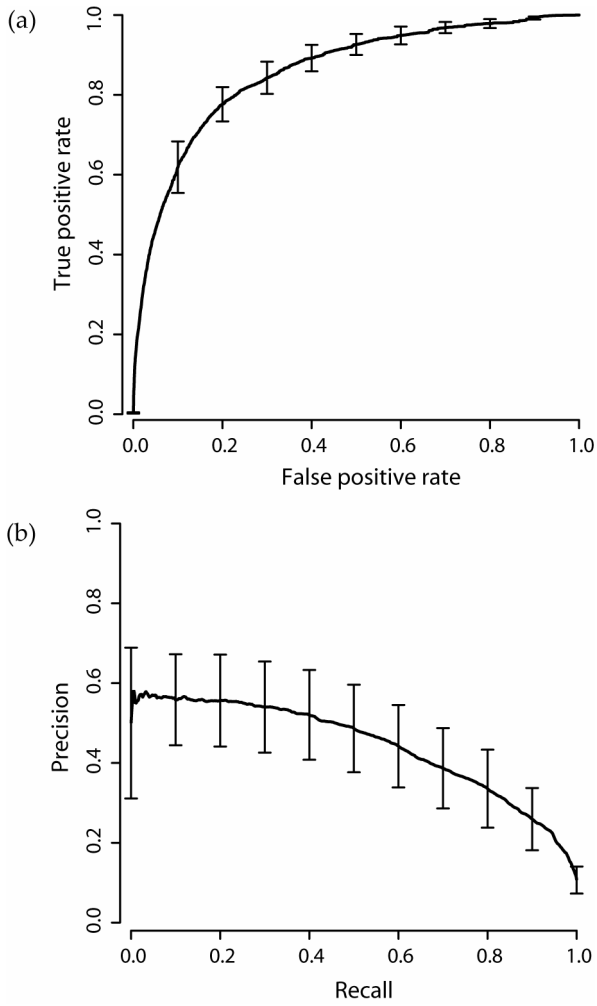

Figure 8-8 (a) Mean ROC and (b) PR curve for the lapse detector based on spectral power (SP) features. The vertical bars indicate standard error on both plots.

Table 8-2 The mean detector performances (ϕ) for systems trained to detect lapses using spectral power, normalized spectral power, and power ratio measures. The detector performances for uniform and constrained least-squares weighting regimes are shown.

Detector features	Detector performance (ϕ)	
	Uniform weights Mean±SE (min, max)	Constrained least-squares weights Mean±SE (min, max)
Spectral power (SP)	0.38±0.06 (0.06,0.59)	0.39±0.06 (0.13, 0.62)
Normalized spectral power (NSP)	0.32±0.05 (0.12,0.49)	0.33±0.05 (0.11,0.57)
Power ratios (PR)	0.34±0.05 (0.12,0.47)	0.33±0.05 (0.10,0.52)
SP+NSP	0.37±0.06 (0.11,0.56)	0.39±0.06 (0.12,0.62)
SP+PR	0.37±0.06 (0.09,0.57)	0.37±0.06 (0.12,0.57)
NSP+PR	0.32±0.05 (0.10,0.50)	0.32±0.05 (0.09,0.49)
SP+NSP+PR	0.36±0.06 (0.10,0.56)	0.36±0.06 (0.11,0.59)
Spectral asymmetry	0.18±0.05 (0.02,0.36)	0.17±0.05 (0.02,0.36)
Spectral coherence	0.15±0.03 (0.00,0.30)	0.15±0.03 (0.00,0.26)

Table 8-3: AUC-ROC and AUC-PR curves for spectral detectors used to detect lapses for both uniform and constrained least-squares weighting regimes. These curves indicate detector performance independent of meta-learner output threshold (cf. Table 8-2).

Detector features	Detector performance			
	Uniform weights		Constrained least-squares weights	
	AUC-ROC (Mean±SE)	AUC-PR (Mean±SE)	AUC-ROC (Mean±SE)	AUC-PR (Mean±SE)
Spectral power (SP)	0.86±0.03	0.41±0.09	0.86±0.03	0.43±0.09
Normalized spectral power (NSP)	0.82±0.04	0.40±0.09	0.82±0.04	0.38±0.09
Power ratios (PR)	0.83±0.03	0.38±0.09	0.83±0.03	0.39±0.09
SP+NSP	0.86±0.03	0.42±0.09	0.86±0.03	0.44±0.10
SP+PR	0.85±0.03	0.42±0.10	0.85±0.03	0.43±0.10
NSP+PR	0.81±0.04	0.39±0.09	0.81±0.04	0.39±0.09
SP+NSP+PR	0.85±0.03	0.42±0.10	0.85±0.03	0.43±0.10
Spectral asymmetry	0.69±0.05	0.24±0.07	0.69±0.05	0.25±0.07
Spectral coherence	0.69±0.03	0.20±0.05	0.69±0.03	0.20±0.06

Figure 8-10 shows the fluctuation in performance with number of PCs in terms of mean phi correlation, AUC-ROC, and AUC-PR for a spectral power based lapse detector. The results show that detector performance reached a plateau after approximately 50 PCs and that adding additional PCs to the model did not cause over-fitting and reduce the overall performance. Therefore, it was decided to include all PCs in the construction of subsequent lapse detector models as this avoids the added complication of determining the number of PCs to be used for model formation.

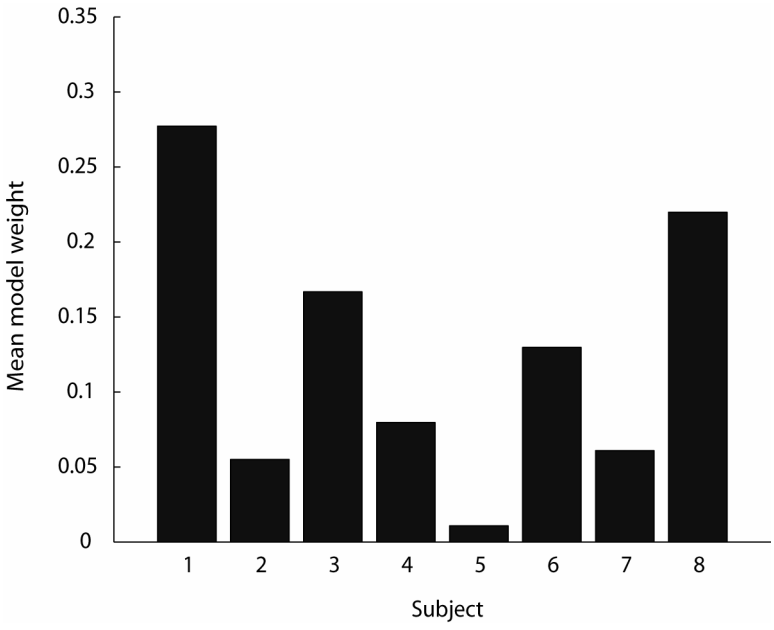

Figure 8-9 Mean meta-learner weights (normalized) of the spectral-power-based lapse detector.

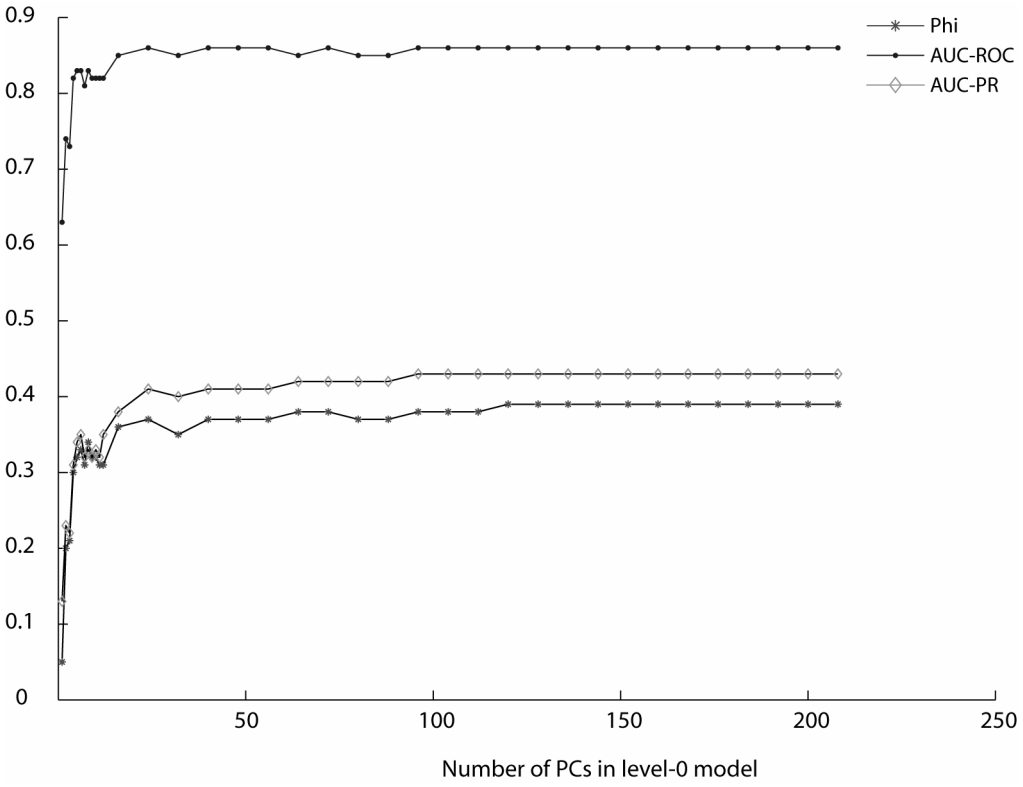

Figure 8-10 Variation in spectral-power-based lapse detector performance in terms of phi correlation, AUC-ROC, and AUC-PR with the number of principal components used to form the level-0 classification models. Note that the values shown are the mean over 8 validation subjects, and the meta-learner utilized constrained-LS weighting.

8.10.2 Simple SP detector model

The performance of a single detection model (cf. stacked model consisting of seven level-0 models followed by a level-1 meta-learner) created by lumping data of seven subjects together and validated using the remaining subject's data was investigated to compare its performance with the stacked approach.

Firstly, one subject was left out for validation. The data of the remaining seven subjects were concatenated and a single classification model created. The lumped data were then fed through the model to determine the optimal output threshold. Finally, the validation subject's data was fed through the model and the phi correlation between the classification model output and the subject's LI calculated. This procedure was repeated until all eight subjects had been used for validation. The mean performance of the simple detector model was the mean over the eight validation runs.

The performance (in terms of phi) for the simple detector model using SP features was 0.31 ± 0.07 (0.06, 0.58). In comparison, when using the same spectral features, the stacked approach with LS-constraints yields a phi value of 0.39 ± 0.06 (0.13, 0.62).

8.10.3 Detector performance – complexity measures only

Table 8-4 and Table 8-5 provide a summary of system performance for a lapse detector using measures of complexity of the EEG. The detector based on LZ complexity measure provided the largest detector performance, as shown by the mean ϕ correlation. Another interesting observation from the results in this table is that a detector using uniform classifier outputs to generate the overall prediction performs better than a detector applying constrained least-squares weights to the classification models to arrive at the final prediction.

8.10.4 Detector performance – spectral and complexity features combined

Since LZ complexity yielded the best detector performance out of the 3 complexity measures, an investigation was performed to evaluate if the detector performance improved if LZ complexity was added to the spectral features. However, as Table 8-6 and Table 8-7 show, no performance improvement was visible after the addition of LZC features to the spectral-power-based detector.

Table 8-4 Mean detector performances (ϕ) for systems trained to detect lapses using FD, ApEn, and LZ complexity measures. The detector performances for uniform and constrained-LS weighting regimes are shown.

Detector features	Detector performance (ϕ)	
	Uniform weights Mean±SE (min, max)	Constrained least-squares weights Mean±SE (min, max)
Fractal dimension (FD)	0.21±0.04 (0.08,0.40)	0.20±0.03 (0.10, 0.38)
Approximate entropy (ApEn)	0.24±0.06 (0.01,0.42)	0.22±0.04 (0.05,0.38)
Lempel-Ziv complexity (LZ)	0.28±0.06 (0.04,0.46)	0.26±0.05 (0.07,0.49)

Table 8-5 AUC-ROC and AUC-PR curves for FD, ApEn, and LZ complexity-based detectors used to detect lapses. These curves indicate detector performance independent of meta-learner output threshold (cf. Table 8-4).

Detector features	Detector performance			
	Uniform weights		Constrained least-squares weights	
	AUC-ROC (Mean±SE)	AUC-PR (Mean±SE)	AUC-ROC (Mean±SE)	AUC-PR (Mean±SE)
Fractal dimension (FD)	0.77±0.03	0.28±0.07	0.75±0.03	0.22±0.05
Approximate entropy (ApEn)	0.77±0.04	0.29±0.07	0.74±0.04	0.23±0.05
Lempel-Ziv complexity (LZ)	0.80±0.04	0.34±0.08	0.78±0.04	0.30±0.08

Table 8-6 Mean detector performance (ϕ) for systems trained to detect lapses using the best spectral features (SP, SP+NSP) with LZ complexity. The detector performances for uniform and constrained-LS weighting regimes are shown.

Detector features	Detector performance (ϕ)	
	Uniform weights Mean±SE (min, max)	Constrained least-squares weights Mean±SE (min, max)
SP+LZC	0.38±0.06 (0.08,0.59)	0.39±0.06 (0.12, 0.63)
SP+NSP+LZC	0.36±0.06 (0.09,0.58)	0.38±0.06 (0.12,0.62)

Table 8-7 AUC-ROC and AUC-PR curves for SP+LZC and SP+NSP+LZC detectors. These curves indicate detector performance independent of meta-learner output threshold (cf. Table 8-6).

Detector features	Detector performance			
	Uniform weights		Constrained least-squares weights	
	AUC-ROC (Mean±SE)	AUC-PR (Mean±SE)	AUC-ROC (Mean±SE)	AUC-PR (Mean±SE)
SP+LZC	0.85±0.03	0.42±0.10	0.86±0.03	0.44±0.10
SP+NSP+LZC	0.85±0.03	0.42±0.10	0.86±0.03	0.44±0.10

8.10.5 Contribution of features to discrimination in best lapse detector

The relative contribution of each EEG feature to the lapse detection model was investigated using the procedure outlined in § 8.9. The detector model based on spectral power was selected for analysis as it displayed the highest performance level ($\phi \approx 0.39$). The proportion of contribution of each spectral feature towards the overall lapse detection model is shown in Figure 8-11. This proportion was determined by summing the contributions of the selected feature across all EEG derivations in all level-0 models. Likewise, the proportion of contribution to the overall detection model by each EEG derivation was calculated by summing the contributions of all spectral features of each EEG derivation across all level-0 models. The proportion of contribution by each EEG derivation is shown in Figure 8-12. Generally, no strong spatial patterns are visible across derivations (apart from T6-O2 which has the largest contribution) indicating that each derivation contributes approximately equally to the overall model. In terms of power spectral features, the spectral power in the alpha band seems to be the largest contributor towards the detection model, as seen in Figure 8-11.

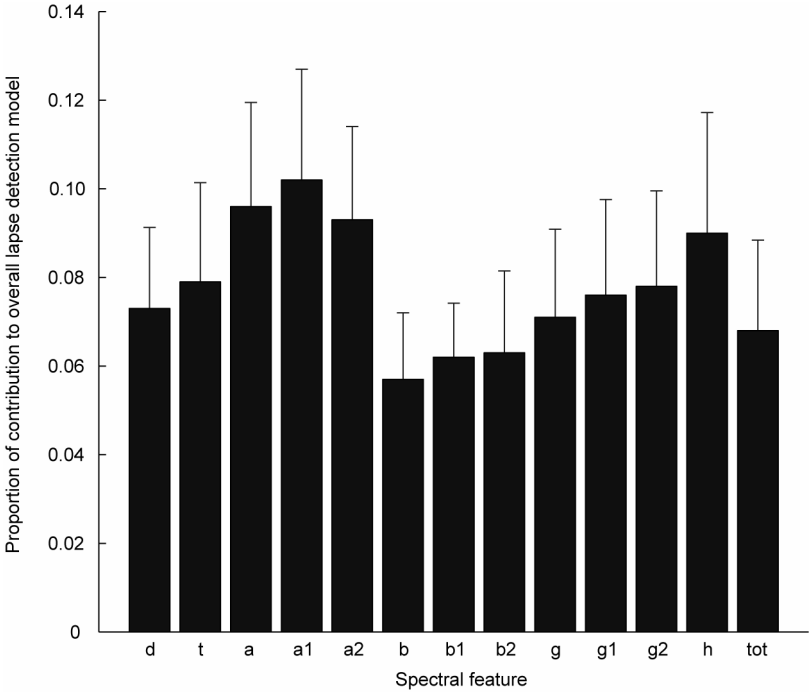

Figure 8-11 Mean proportion of contribution by each spectral feature to the overall lapse detection model. The contribution of each spectral feature was found by summing the contributions of the selected feature across all EEG derivations. The spectral features were delta (d), theta (t), alpha (a), alpha 1 (a1), alpha 2 (a2), beta (b), beta 1 (b1), beta 2 (b2), gamma (g), gamma 1 (g1), gamma 2 (g2), high (h), and total (tot).

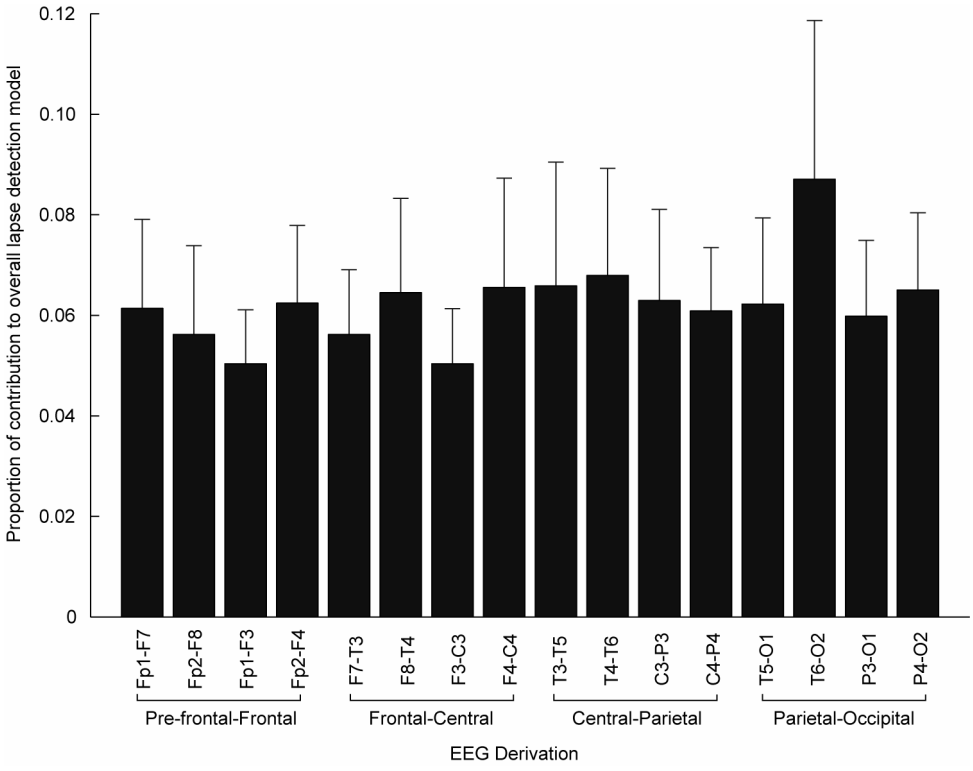

Figure 8-12 Mean proportion of contribution by each EEG derivation to the overall lapse detection model (based on power spectral features). The contribution of each derivation was found by summing the individual contributions of all spectral features within each derivation.

8.10.6 Tonic *vs*. phasic changes

An analysis was carried out to determine if the lapse detector was predominantly detecting transient or phasic events (lapses of duration of few seconds) or tonic changes in the level of drowsiness which typically occur on a sub-minute to minute scale. To investigate this, a bi-directional, 4th-order Butterworth filter (low-pass) was used to smooth the output of the meta-learner. As before, an optimal threshold value for the smoothed output was determined during the training phase and applied during the validation phase to obtain the overall prediction. This analysis was restricted to the lapse detector that yielded the highest performance (i.e., spectral-power-based detector).

Figure 8-13 shows the detector performance in terms of mean phi, AUC-PR, and AUC-ROC as filter cut-off is varied. The filter cut-off frequency f_c is related to the filter time constant τ by

$$\tau = \frac{1}{2\pi f_c}. \tag{8-15}$$

Peak detector performance occurs at a filter cut-off frequency of approximately 0.1 Hz ($\tau = 1.6$ s) as seen in Figure 8-13. Furthermore, a steep drop in detector performance is observable for filter cut-off values less than approximately 0.05 Hz ($\tau = 3.2$ s). Detector performance also gradually decreases as the filter cut-off is increased, say, beyond 0.15 Hz ($\tau = 0.9$ s).

Since peak detector performance occurred at a filter time constant of approximately 1.6 s, one can conclude that the lapse detection system is predominantly detecting phasic events rather than tonic changes in drowsiness. If the system was predominantly detecting tonic changes in the level of arousal, one would have expected the peak detector performance to occur at a much lower cut-off value than 0.1 Hz (say 0.005 Hz).

8.10.7 Detection of lapse events

Up to this point, this chapter has assessed the performance of lapse detectors in terms of their ability to detect the lapse state in 1-s epochs. The goal of this analysis was to determine the detector's ability to detect discrete *lapse events*. This analysis was restricted to the spectral-power-based detector.

The following procedure was used to determine lapse event detection performance:

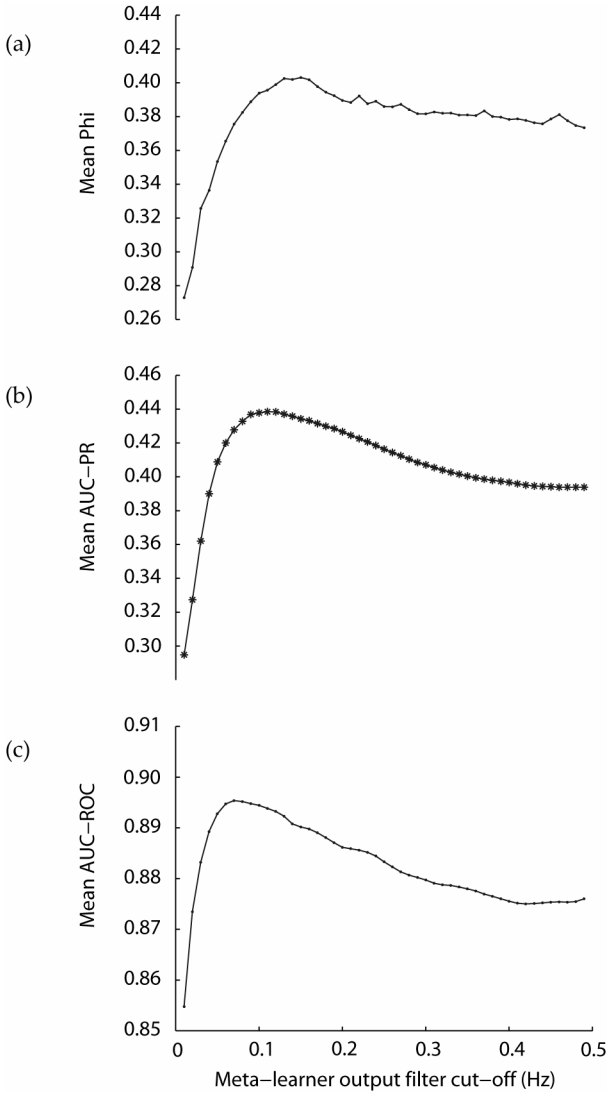

Figure 8-13 Effect of filter cut-off value applied to the meta-learner output *vs.* lapse detector performance in terms of (a) mean phi, (b) AUC-PR, and (c) AUC-ROC. Results are shown for the best performing spectral lapse detector.

1. An *event signal* was created. This was the same length as the LI with a sampling frequency of 1 Hz.
2. The event signal was initialized to a default value of 0, which was defined as a true negative (TN) event.
3. A pre-determined optimum threshold (see § 8.8) was applied to the overall lapse detector output to obtain binary *detector lapse events*. A detector output of 1 was defined to correspond to a *detector lapse event*, and a detector output of 0 to correspond to the responsive state.
4. The gold standard (i.e., LI) was traversed until a *gold standard lapse event* was encountered.

5. The detector output was checked to see if it equaled 1 during any portion of the gold standard lapse event. If yes, the entire portion of the event signal corresponding with the gold standard lapse event was marked as a true positive (TP) event. Furthermore, the TP event was extended at either end to include any overlapping detector lapse event. If the detector output was 0 during the entirety of the gold lapse event, the corresponding region of the event signal was marked as a false negative (FN) event.
6. The detector output was traversed until a detector lapse event was encountered.
7. The region of the event signal corresponding to the detector lapse event was checked. If marked as a TN, the region in the event signal corresponding with the detector lapse event was re-marked as a false positive (FP) event.
8. The number of TPs, FPs, TNs, and FNs in the event signal were counted and the following performance parameters calculated:

    Sensitivity = TP / (TP + FN)

    Selectivity = TP / (TP + FP)

    Specificity = TN / (TN + FP)

    Negative predictive value = TN / (TN + FN)

    Accuracy = (TP + TN) / (TP + TN + FP + FN)

Note that sensitivity is also referred to as the *true positive rate* or *hit rate*. Selectivity is referred to as the *positive predictive value* or *precision*. All parameter values are expressed as percentages.

An example of how the event signal was generated from gold standard lapse events and detector lapse events is shown in Figure 8-14. The performance of the spectral-power-based lapse detector, in terms of its ability to detect lapse events is summarized in Table 8-8. Overall event detection performance was calculated by concatenating the data from all 8 subjects and yielded an overall sensitivity of 73.5%, selectivity of 25.5%, and an accuracy of 61.2%.

Note that the total number of lapses listed in the 2nd column of Table 8-8 for each subject differs from the lapse counts present in Chapter 7. The reason for this apparent discrepancy is due to the decision to exclude data contaminated with artifacts from being used in lapse detector model formation (see § 8.2). This process resulted in the exclusion of approximately 8.5% of the total number (57600) of available epochs from the 8 subjects.

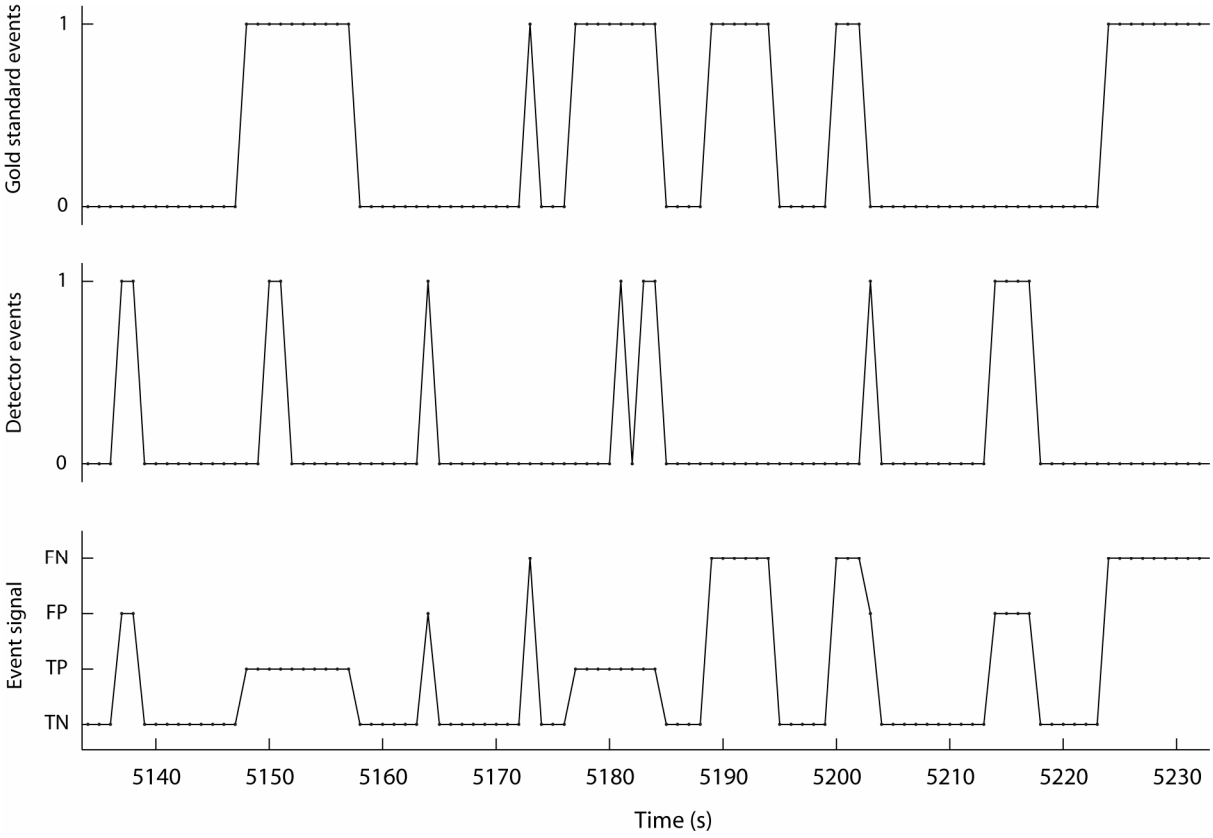

Figure 8-14 An example illustrating how the Event Signal (bottom plot) was derived from gold standard events (top) and detector events (middle). Note that in the top two plots, values of 1 and 0 indicate lapse events and responsive states, respectively.

Table 8-8 Event detection performance of the spectral-power-based lapse detector in terms of TPs, TNs, FPs, FNs, and sensitivity, specificity, selectivity, NPV, and accuracy percentages.

		Event detector performance								
Subject	**Total lapses**	**TP**	**TN**	**FP**	**FN**	**Sen.**	**Spec.**	**Sel.**	**NPV**	**Accy**
1	235	118	328	102	117	50.2	76.3	53.6	73.7	67.1
2	77	59	289	221	18	76.6	56.7	21.1	94.1	59.3
3	12	9	299	290	3	75.0	50.8	3.0	99.0	51.2
4	55	52	483	433	3	94.5	52.7	10.7	99.4	55.1
5	46	45	292	246	1	97.8	54.3	15.4	99.7	57.7
6	109	70	327	223	39	64.2	59.4	23.9	89.3	60.2
7	194	155	453	278	39	79.9	62.0	35.8	92.1	65.7
8	189	166	352	179	23	87.8	66.3	48.1	93.9	72.0
Overall	**917**	**674**	**2823**	**1972**	**243**	**73.5**	**58.9**	**25.5**	**92.1**	**61.2**

8.10.8 Effect of lapse duration on detection

An analysis was conducted to determine the relationship between the duration of a lapse and the likelihood of it being detected. As before, the best-performing lapse detector (i.e., the spectral-power-based detector) was used.

Figure 8-15 shows the distribution of detected, missed, and false positive lapses with duration. These histograms were generated by pooling data from all eight subjects. The system successfully detected 362 lapses and missed 621 lapses over the eight subjects. The median durations of detected and missed lapses in the pooled data were 4.0 s and 3.0 s, respectively (note that the lapse detection resolution is 1.0 s), the difference being marginal (Wilcoxon: $p = 0.0587$). There were 424 false positive detections over the eight subjects.

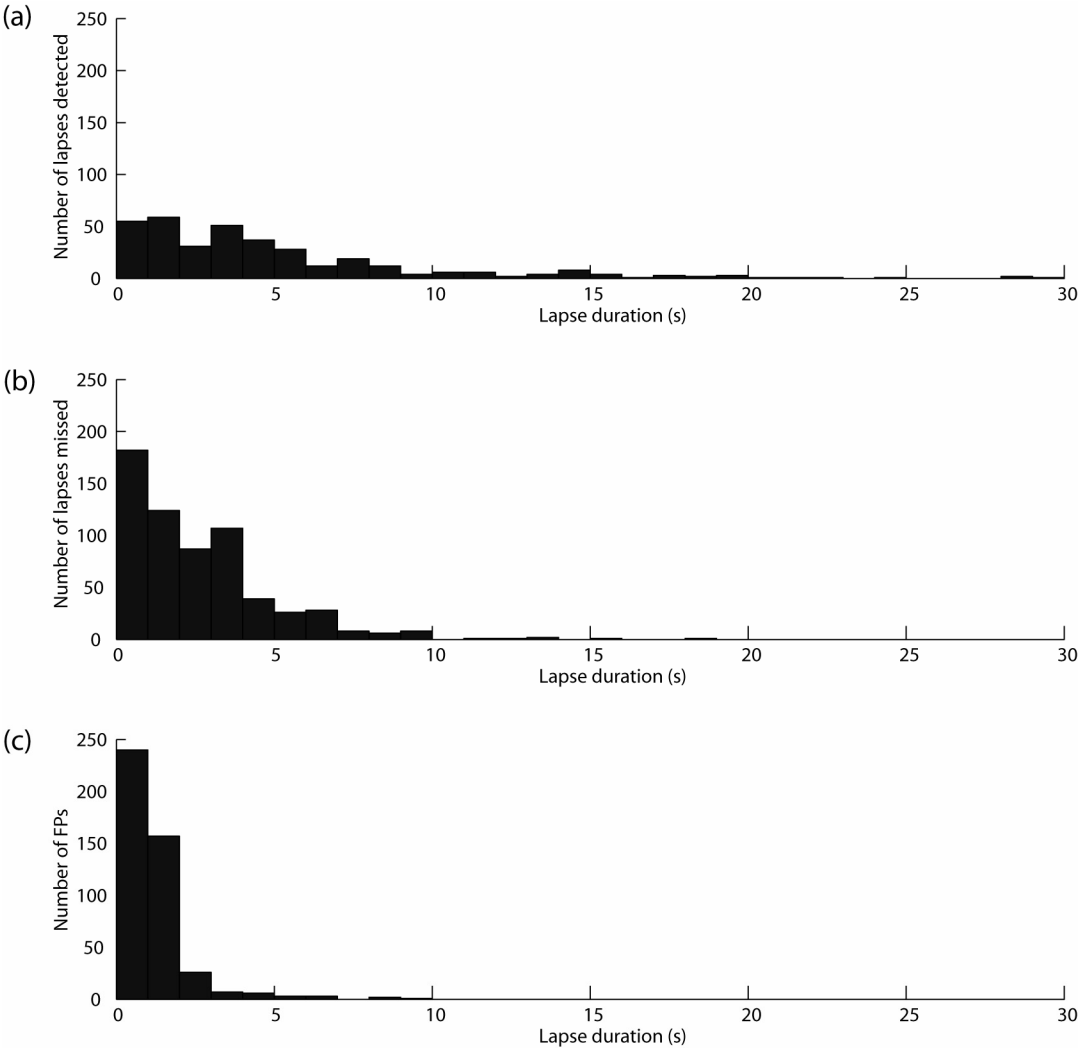

Figure 8-15 Histograms depicting lapse duration against (a) number of lapses detected (b) number of lapses missed, and (c) false positive lapses. For clarity, the maximum lapse duration depicted is limited to 30 s. There were eight lapses longer than 30 s (all detected) and one false positive detection (37 s).

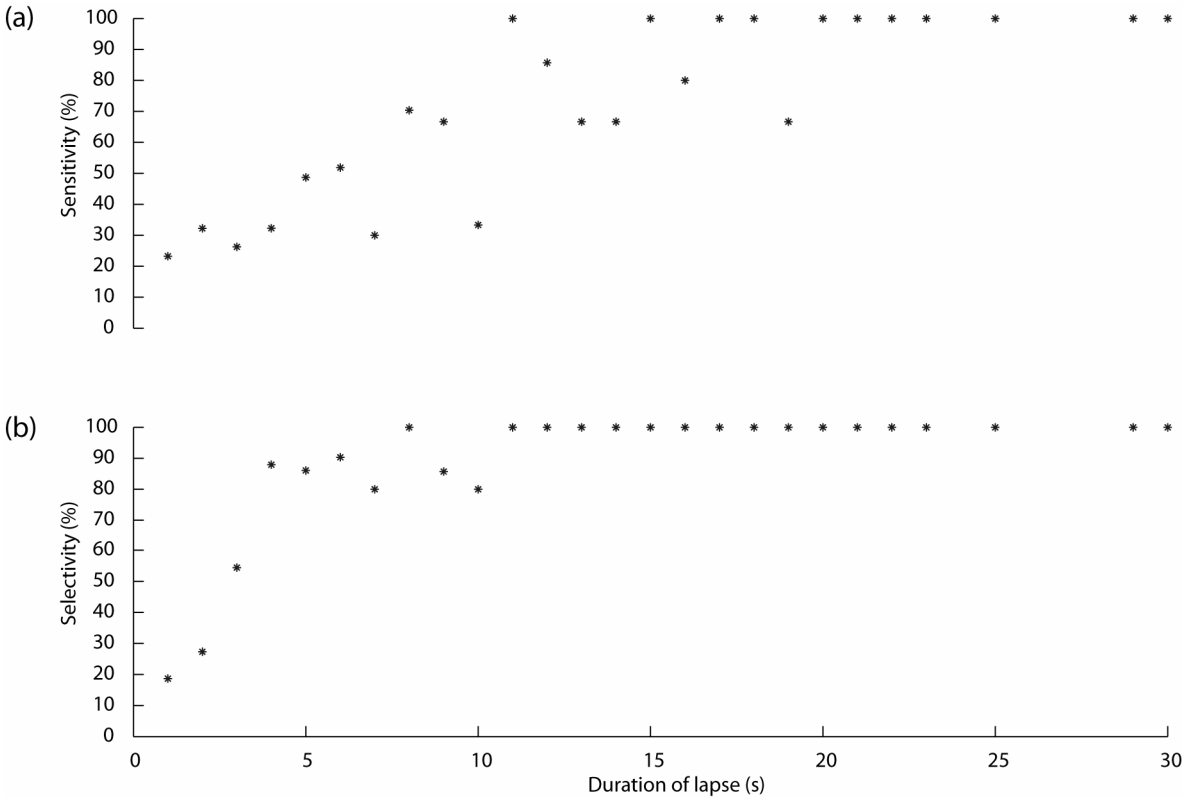

Figure 8-16 Changes in (a) sensitivity and (b) selectivity of the spectral-power-based lapse detector with lapse durations less than 30 s. Note that sensitivity = TP / (TP + FN) and selectivity = TP / (TP + FP). Since the data set did not contain lapses of all durations, there are certain durations for which it was not possible to calculate sensitivity and/or selectivity.

Sensitivity and selectivity of the lapse detection system with lapse duration is shown in Figure 8-16. An improvement in detector sensitivity and selectivity with increasing lapse duration can be seen. As shown in Figure 8-16 (a) all lapses greater than 20 s were successfully detected by the system. There was only one false positive detection beyond 10 s (duration = 37 s).

8.11 Discussion

This chapter described the procedure used to investigate the detection of lapses from the EEG. Firstly, an overview of the signal processing methods used to extract linear and non-linear features of the EEG was provided. This was followed by a description of how the extracted EEG features were used to create models capable of detecting lapses from the EEG. Finally, results for the various detection models were presented and the performance of the different models compared.

Results were presented for lapse detectors based on both linear (spectral), non-linear, and a combination of both linear and non-linear features of the EEG. This showed that the best

detector performance (in terms of the highest mean ϕ) was achieved using the detector model created using SP features with constrained least-squares used to determine the meta-learner weights. This model showed a mean correlation of $\phi \approx 0.39 \pm 0.06$ (AUC-ROC = 0.86 ± 0.03; AUC-PR = 0.43 ± 0.09). Classification models created using NSP or PR features (in conjunction with the use of constrained least-squares to determine meta-learner weights) had lower mean performances than the SP-based detector ($\phi = 0.33 \pm 0.05$ for both cases). The performance of NSP/PR detectors showed a marginal increase when SP features were added to the NSP/PR features (see Table 8-2). In contrast, the mean performance of lapse detectors based on spectral asymmetry and coherence was less than half the performance shown by SP/NSP/PR feature-based detectors. This suggests that only a sparse amount of information useful for detecting lapses is contained in the spectral asymmetry and coherence features of the EEG. The case of combining spectral asymmetry and coherence with spectral power features to form a detector could not be pursued as there was insufficient computer memory to process such a large feature set.

The LZ complexity feature-based detector showed the highest performance ($\phi = 0.28 \pm 0.06$; AUC-ROC = 0.86 ± 0.03; AUC-PR = 0.43 ± 0.09) of the 3 non-linear feature-based detectors, followed by the detector based on approximate entropy. Interestingly, the use of uniform meta-learner weights showed a marginally higher performance than using constrained LS meta-learner weights (cf. spectral detectors) for all 3 non-linear detectors in terms of ϕ, AUC-ROC, and AUC-PR.

The performance of a lapse detector created by adding the best performing non-linear features (LZ complexity) to the best linear features (SP) was no greater than the detector based on SP alone. This suggests that non-linear features contribute no additional information to the detector and that they effectively contain information similar to SP features. Note that LZ complexity generated 1 feature per channel to give 16 features per subject (1 feature/channel × 16 channels) whereas SP contained (13 features/channel × 16 channels) 208 features. One might therefore propose that LZ features have a disadvantage over SP in that they use a much smaller feature set (16 features) compared to SP, resulting in a detector with performance inferior to an SP-based detector. However, analysis showed (Figure 8-10) that an SP detector limited to using the first 16 PCs still performed better than the LZ detector ($\phi = 0.36 \pm 0.05$ *vs.* 0.28 ± 0.06; AUC-ROC = 0.85 ± 0.03 *vs.* 0.80 ± 0.04; AUC-PR = 0.39 ± 0.09 *vs.* 0.34 ± 0.08).

Figure 8-10 shows the performance of the SP-based lapse detector *vs.* the number of PCs used to form the classification model. It shows that performance initially increased rapidly with the increasing number of PCs and reached a plateau after approximately 50 PCs. There was no substantial drop in performance as the number of PCs used to form the model

increased. This suggests that adding PCs after performance has peaked did not cause over-fitting in this case. Presumably this may be due to level-0 models assigning very small weights to the PCs that did not contribute to classification, thereby mitigating the likelihood of over-fitting occurring. Even though this may be the case, it is best to use as few PCs as possible to reduce the computational effort. The fact that detector performance did not cause over-fitting when all PCs were used suggests that one could have also used the raw features (instead of PCs) to form the detection models and expect similar results.

Results showed that the performance of a spectral-power based lapse *event* detector is moderate at best with an accuracy of 61.2%. The system detected a very large number of FPs which resulted in low selectivity. Selectivity could have been improved by increasing the output threshold of the overall detector, although this would be at the expense of decreased sensitivity. In general, since the subject is not lapsing for the majority of time, it is not surprising to observe a large number of TN events, which tends to bias the detector towards high specificity, NPV, and accuracy values. However, it is important to acknowledge that such seemingly good values can be obtained by chance alone, and that sensitivity and selectivity values are more reliable in this context.

It was also shown that the lapse detection system is more likely to detect longer lapses indicated by increasing values of sensitivity and selectivity with lapse duration. This is presumed to be primarily due to longer lapses having more pronounced EEG spectral changes related to microsleeps than shorter lapses. However, an increase in detection sensitivity with increased duration of lapse could also occur due simply to chance (i.e., even if the detector output had no relationship with the occurrence of lapses) although this would also tend to be offset by a concomitant increase in longer-duration false detections.

As mentioned earlier in the chapter, EEG epochs with z-scores >30 were excluded as artifacts, and were not used for training and testing the lapse detection models. This may have biased the detector towards a better level of performance. However, because the excluded proportion of epochs was relatively low (8.5%), it is unlikely that their elimination caused the system to display a substantially higher level of performance than what would have been obtained if 'contaminated' data was used to test the models.

Overall, the levels of performance and reliability demonstrated by the lapse detection models are not sufficient for real-time detection. However, they are encouraging since this task involved the added challenge of detecting lapses on a 1-s resolution. This fact was confirmed in § 8.10.6 which demonstrated that optimum detector performance occurred at a filter cut-off of 0.1 Hz, corresponding to a filter time constant of approximately 1.6 s, indicating that the system was detecting phasic changes. This contrasts with studies that used a larger time-

scale to smooth performance metrics resulting in an estimate of alertness/drowsiness on a scale of 1-min or more (Jung *et al.*, 1997; Lin *et al.*, 2006; Lin *et al.*, 2005a; Lin *et al.*, 2005b; Makeig *et al.*, 1996).

Makeig *et al*. (1996) used a feed-forward neural network to estimate changes in alertness via the EEG cross-spectrum in a group of ten subjects performing a 30-min auditory detection task. The local error rate was smoothed using a 95-s window. Individual models were formed to estimate alertness using the first session of the subjects. The generalizability of the model was tested using data from the second session. The performance of the approach, measured in terms of the mean correlation coefficient between actual and estimated error was 0.67.

The model proposed by Jung *et al*. (1997) (described in § 4.4.1) to estimate task performance error in an auditory detection task yielded a minimum mean r.m.s. error of 0.16 ± 0.05 when trained on one session and tested on a 2nd session from the same subject. The time resolution of the performance estimate was limited to minute-scale as a 93.4-s window was used to smooth the performance error. It is also important to note that Jung *et al*.'s approach required detection models to be created for individual subjects.

Similarly, the approaches of Lin *et al*. (2006; 2005a; 2005b) (described in § 4.4.1) to estimate changes in driving performance via EEG power spectrum also had approximately minute-scale time resolution due to the 90-s window used to smooth the driving performance index. Their first approach (Lin *et al.*, 2005a), which utilized the log sub-band power spectrum, correlation analysis, PCA, and linear regression models yielded a between-session performance of 0.53 ± 0.12 (mean correlation coefficient). The next approach (Lin *et al.*, 2005b), which used EEG-spectral power, ICA, correlation analysis, and a linear regression model displayed a mean correlation coefficient of 0.88 ± 0.05 for between-session testing. The third approach (Lin *et al.*, 2006) utilized an adaptive feature selection mechanism, based on correlation coefficients between log band power of the ICA components and the driving performance index, and an ICA-mixture-based fuzzy neural network to estimate driving performance from the EEG. The mean correlation between the model output and the performance index for between-session testing was 0.91 ± 0.03. It is expected that Lin *et al.*'s approaches would yield much lower mean performance had they attempted to detect lapses at a higher temporal resolution (e.g., 1-s scale).

One of the main considerations taken to develop the current lapse detection system was the ability to generalize well to new subjects. This is another important distinction between the approach outlined in this chapter and the approaches mentioned above (Jung *et al.*, 1997; Lin *et al.*, 2006; Lin *et al.*, 2005a; Lin *et al.*, 2005b) which required training a model for each

individual, to predict their performance in subsequent sessions. Since a model tuned for each subject takes subtle differences in the individual's EEG into consideration, it is likely to yield superior performance. However, a major disadvantage is that each user requires extensive training prior to use of the device.

A member of the Christchurch Lapse Research Programme used normalized EEG log-power spectrum inputs to train a long short-term memory recurrent neural network (RNN) based lapse detector which demonstrated an overall performance level of AUC-ROC=0.84 ± 0.02; AUC-PR = 0.41 ± 0.08; mean phi = 0.38 ± 0.05 (Davidson *et al.*, 2007). The RNN approach was applied to the same dataset described in this chapter. The RNN approach and the current linear approach used a time resolution criteria of 1.0 s for detection of lapses. The results presented in this chapter demonstrate that a relatively simple linear approach based upon spectral power is capable of achieving a very similar level of performance to the RNN approach presented by Davidson *et al*. The linear approach has the added advantage of being computationally less intensive than the RNN to train. The comparable level of performance between the RNN and the linear spectral power approach described in this work is somewhat surprising and suggests that the linear detector has superior parsimony. It also suggests that lapses involve, at most, only a mild non-linearity as otherwise one would expect neural-network-based detectors and linear-nonlinear- feature detectors (i.e., § 8.10.4) to achieve higher performances if the solution was substantially non-linear (Bishop, 1995).

The fact that constrained-LS weights gave no clear improvement in detection performance over uniform weights for the meta-learner was surprising. Stacked generalization was used to combine the model outputs because this was expected to be the best method of combining the level-0 models by determining the optimal weights for the level-0 models using the training data. However, only a slight trend in increased performance (in terms of mean phi) was observed in lapse detectors based on SP features. In fact, non-linear feature based lapse detectors with uniform weights showed a trend towards slightly outperforming the constrained-LS weighted meta-learner. It is important to note that performance levels between the constrained-LS and uniform meta-learner weighted detectors were approximately equal. This suggests that the eight level-0 models contributed approximately equally to the overall output. However, this can be dismissed by observing Figure 8-9 and noting that the mean constrained-LS weights for the eight SP-based detectors are not uniform. A possible reason for the lack of substantial improvement in detector performance with the use of the stacked approach may be due to the level-0 models being very similar to each other. It has also been suggested that one must use dissimilar predictors to obtain the most improvement in performance when using a stacked system (Brieman, 1996b).

The advantage of creating multiple models and the resulting increase in detector performance was revealed in § 8.10.2. Using eight level-0 models and a meta-learner, resulted in a mean phi correlation of 0.39 for the SP detector, whereas lumping all the features from seven subjects to create a single model and validating the lumped model on the eighth subject resulted in a mean phi of 0.31. The lumping of data prior to model formation may result in the model being biased towards certain subjects, such as ones with the most lapses, resulting in a loss of generalizability. Using a stacked approach yields better performance as the level-0 models are adjusted by the meta-learner according to how well they generalize over the training set. It is expected that the mean detector performance would increase if the size of the training set was increased. However, this was not possible in this work as the dataset was limited to eight subjects.

Investigation into the features and channels that contributed most to the SP-based lapse detector (which showed the highest level of performance out of all the detectors) revealed that the alpha band contributed the most to the overall detection model. This result fits with previous research which has demonstrated relatively low correlations between amplitude changes in the alpha range during or immediately after auditory lapses (Makeig and Jung, 1996) and visual lapses (Cajochen *et al.*, 1999). A decrease in alpha power – as theta power increases – has also been reported during microsleeps (Harrison and Horne, 1996; Valley and Broughton, 1983). It is possible that this decrease in alpha was 'selected' by the lapse detector as a useful cue to determine the occurrence of lapses. Torsvall and Åkerstedt (1988) noted that alpha power density peaked during a ~22 s period preceding the onset of a 'dozing off' event during a visual tracking task.

It was surprising to find that that theta power, shown to have one of the highest (albeit small) correlations with lapses in Chapter 7, and also associated in the literature with reduced auditory alertness (Huang *et al.*, 2001; Jung *et al.*, 1997), driver fatigue (Lal and Craig, 2005), and microsleeps (Harrison and Horne, 1996; Priest *et al.*, 2001; Valley and Broughton, 1983), does not substantially contribute to the overall spectral-power-based lapse detector. This suggests that changes in theta power did not feature prominently in this dataset. Analysis showed that beta power contributed the least to the SP detector of all the features, contrasting with Belyavin and Wright (1987) who found beta power to be the most useful discriminator of worsening vigilance in a visual vigilance and letter discrimination task. It is possible that this apparent discrepancy may be due to beta power being correlated with the depth of drowsiness rather than lapses.

Cajochen *et al.* (1999) reported that frontal EEG activity (1–7 Hz) showed an increase with deteriorating performance in sleep-deprived subjects and Santamaria and Chiappa (1987) reported an increase in centrofrontal alpha, often occurring concurrently with a decrease in

amplitude of occipital alpha, as drowsiness increased. However, results from this work do not support this as there was approximately an equal amount of contribution towards the overall detection model from all EEG derivations, indicating that there is no strong spatial pattern that could be used to detect lapses (apart from derivation T6-O2 which showed a marginally higher contribution). Again, the apparent discrepancy may be due to the aforementioned studies correlating with the depth of drowsiness rather than lapses.

A possible reason for the overall low detector performance may be the fact that a majority of lapses required to be detected were only a few seconds in duration. These short lapses may have been of insufficient duration to cause substantial changes in the EEG which could be extracted via various signal processing methods.

Due to the type of tracking task used, it was not possible to use tracking error alone to determine a subject's performance. This was due to the tracking error becoming low purely due to chance when the target waveform moved closer to the response cursor. However, in combination with independently rated facial video, it was possible to achieve an approximately 1-s time resolution for the gold standard.

CHAPTER 9

Conclusions and Future Research

9.1 Key findings and discussion

The major motivation behind this research project was to identify reliable physiological cues indicative of lapses [particularly behavioural microsleeps (BMs)] from the EEG, which could be used in the development of a real-time lapse detection (or, better still, prediction) system capable of continuously monitoring an individual and providing warnings of impending lapses. Additionally, the project also focused on contributing towards an increased understanding of the characteristics of lapses in responsiveness in normal subjects.

The first study showed that detection of lapses by human expert raters was poor, with only 6 of 101 lapses being detected while viewing the EEG and electro-oculogram (EOG) simultaneously, and with inter-rater agreement on only 2 lapses. Slightly more lapses were detected viewing EEG alone and none were detected when viewing EOG alone. Several factors such as the lack of a continuous measure of performance, the short (10 min) test duration, the psychomotor vigilance task (PVT) task placing a heavy cognitive demand on the subject (and consequently temporarily increasing performance), lack of an independent measure of alertness such as video, lack of symmetric EEG channels, lack of separate horizontal and vertical EOG channels, and different rating criteria may have contributed to the poor correlation between the expert ratings and performance. Alternatively, all of the lapses may have been lapses of sustained/focused attention rather than BMs, and the EEGer may only be able to detect BMs from the EEG (although this has not been determined). Results demonstrated that detecting lapses based on EEG and EOG was not a trivial task for expert EEG raters. This suggested that, unlike an automated sleep stager, an automated lapse detection system needs to identify features that are not readily visible in the EEG.

Limitations in the air traffic controller (ATC) dataset led to a study where more comprehensive physiological and performance data were collected from 15 normal non-sleep-deprived subjects performing a continuous tracking task during normal work hours. Frequent unequivocal lapses in a substantial proportion of subjects were observed. Fourteen of the 15 subjects had one or more lapses[5], with an overall rate of 39.3 ± 12.9 lapses per hour (mean ± SE) and lapse duration of 3.4 ± 0.5 s. The study also showed that lapsing and tracking error increased during the first 30 min of a 1-h session but subsequently decreased over the second 30 min, despite the absence of external temporal cues. Spectral power was found to be higher during lapses in the delta, theta, and alpha bands, and lower in the beta, gamma, and higher bands, but correlations between changes in EEG power and lapses were low. It was concluded that lapses are a frequent phenomenon in normal subjects – even when not sleep-deprived – engaged in an extended monotonous continuous visuomotor task. This is the first study to investigate and report on the characteristics of complete lapses of responsiveness during an extended continuous tracking task (CTT) in non-sleep-deprived subjects (Peiris *et al.*, 2006).

Best lapse detector performance was achieved using a detector based on spectral power (SP) features with constrained least-squares fitting used to determine the meta-learner weights when combining predictions of individual subjects models (mean correlation of $\phi = 0.39 \pm 0.06$). Normalized spectral power (NSP) and power ratio (PR) based detectors showed lower performance levels ($\phi = 0.33 \pm 0.05$ for both cases). The mean performance of lapse detectors based on spectral asymmetry and coherence was less than half the performance shown by SP/NSP/PR feature-based detectors. The Lempel-Ziv complexity feature-based detector showed the highest performance ($\phi = 0.28 \pm 0.06$) out of the three non-linear feature-based detectors, followed by the detector based on approximate entropy. The performance of a lapse detector created by adding the best performing non-linear features (Lempel-Ziv complexity) to the best linear features (spectral power) revealed that combining these features produced no improvement in performance over the detector based on SP alone, suggesting that Lempel-Ziv complexity features contributed no additional information. Analysis also showed that creating multiple lapse detection models and combining them to form an overall detector results in an improvement in performance over creating a single model by concatenating data of all subjects. It was also established that the

[5] In this case, the presence of either a video BM or a tracking flat-spot was counted as a lapse. In contrast, only subjects who had a definite BM (tracking flat-stop coincident with a video BM) were selected to form lapse detection models (*N*=8).

lapse detection model was detecting transient events rather than phasic changes in level of drowsiness. However, the event detection performance of the model was moderate at best.

It was decided to exclude EOG signals from model forming due to time restrictions on the project and therefore only EEG analysis was carried out. The work performed did not yield highly reliable EEG indicators of lapses, and suggested a large degree of intra- and inter-subject variability. No substantial pre-cursors were found which could enable lapses to be predicted. Lempel-Ziv complexity performed best out of the non-linear models but the performance was still relatively poor. It was also determined that there were no specific EEG locations that contributed substantially more in terms of lapse detection than other sites. Of the spectral features, power in the alpha band showed the highest contribution towards the detection model.

9.2 Review of hypotheses

Hypothesis 1: Electroencephalographers (expert EEG readers) are able to identify cues in the EEG and/or EOG which indicate lapses.

Minimal evidence was found to support this hypothesis with only 7 of the lapses being detected from EEG/EOG by, at most, 2 of the 4 EEGers. Although trained scorers are able to classify sleep stages using the EEG and EOG reasonably reliably, the rating study suggests that cues present in the EEG and/or EOG of most (if not all) lapses are too subtle to be detected visually unless, possibly, accompanied by EEG microsleeps occurring during deep drowsiness.

Hypothesis 2: Normal non-sleep-deprived subjects can have multiple behavioural microsleeps of several seconds while performing extended sustained attention tasks.

Strong evidence was found to support this hypothesis with 8 of the 15 subjects having at least one behavioural microsleep during the two 1-hour sessions. Furthermore, 14 of the 15 subjects had one or more definitive lapses, although these were not all necessarily BMs.

Hypothesis 3: Non-linear signal processing techniques can enhance the detection of lapses over that obtainable from linear methods.

No evidence was found to support this hypothesis. Non-linear features of the EEG could not enhance detection performance over that obtainable from linear spectral power-based detection. This suggests that there are no reliable non-linear features in the EEG correlated with lapses.

Hypothesis 4: Lapses in task performance are reflected in decreases in coherence across specific brain sites.

The results provided only minimal support for this hypothesis with a coherence-based lapse detector showing a much lower correlation with lapses to that of spectral-power-based detectors. This suggests that (a) any decreases in coherence during lapses are small and/or inconsistent and (b) changes in intra-cortical coherence cannot be used to aid the detection of lapses.

Hypothesis 5: Lapses in task performance are reflected by increased spectral asymmetry between different EEG sites.

The results provided only minimal support for this hypothesis with a spectral-asymmetry-based lapse detector showing a much lower correlation with lapses to that of spectral-power-based detectors. This suggests that (a) any decreases in spectral asymmetry during lapses are small and/or inconsistent and (b) changes in spectral asymmetry cannot be used to aid the detection of lapses.

9.3 Review of project goals

Goal 1: Conduct a review of the literature to discover and evaluate previous approaches, based upon EEG and/or EOG, used in the detection and/or prediction of drowsiness and lapses.

As presented in Chapter 4, this goal was successfully achieved during the initial stages of the project. The literature review showed that there are only a few approaches that have attempted the detection of lapses (cf. drowsiness estimation, sleep-staging). It also showed that two key aspects poorly addressed to date are (1) characteristics of lapses and (2) methods for detection of lapses, particularly with high temporal resolution.

Goal 2: Find subject-independent features in EEG/EOG which provide reliable indications of lapses and drowsiness.

As the project evolved, a decision was made to focus on detecting lapses via the EEG. Unfortunately, highly reliable indicators to detect and/or predict lapses of responsiveness from the EEG were not found. There is no evidence in the literature of other groups who have succeeded to a substantial degree in this area. Goal two was achieved in terms of determining that alpha and high bands contributed the most to the spectral power-based detector (which admittedly had only modest performance).

Goal 3: Investigate the efficacy of several advanced signal processing techniques as a means of substantially improving the detection accuracy of features and precursors in the brain's electrical activity and/or eye movements, thus reliably detecting drowsiness and lapses.

The detection of lapses of responsiveness from the EEG was attempted via power spectral analysis, spectral coherence and asymmetry, fractal dimension, approximate entropy, and Lempel-Ziv complexity. Best lapse detector performance was achieved using the detector model created using spectral power features. However, the performance of this detector was moderate at best ($\phi \approx 0.39 \pm 0.06$; AUC-ROC = 0.86 ± 0.03; AUC-PR = 0.43 ± 0.09).

Goal 4: Determine the minimum number of EEG and/or EOG channels, and optimal placement of electrodes to achieve a high degree of performance (i.e., eliminate redundant information).

The pursuit of this goal was dependent on the success of goals two and three which, as mentioned earlier, was limited. The main purpose of goal four was to improve real-time lapse detection by minimizing the amount of data collected, consequently improving processing time. However, since highly reliable cues could not be determined, there was no immediate need to pursue this goal.

Goal 5: Confirm the presence and investigate the characteristics of lapses in normal non-sleep-deprived subjects.

Despite some initial apprehensions as to whether subjects would lapse at all during the 1-h tracking task, the continuous tracking study demonstrated that serious lapses can occur in young, healthy, non-sleep-deprived adult males to a much greater extent than previously recognized. Chapter 7 provided a detailed analysis of the characteristics of these lapses.

9.4 Critique

The lack of an additional measure of alertness (such as facial video) in the ATC dataset hindered the researchers' ability to make definitive conclusions about the rating study as there was no way of knowing whether a lapse was caused by loss of sustained/focused attention, a distraction, or a behavioural microsleep, with the first two types presumed to be essentially impossible to detect from the EEG. Although suitable for its original purpose, the duration of the ATC experimental sessions was not long enough to induce lapses in most subjects. Even then, the lapses in the ATC dataset were mostly of short duration and arbitrarily defined as reaction times >500 ms.

The CTT study relied on the subjects informing the investigators about their previous night's sleep pattern. This was to ascertain that they were not sleep deprived. However, no objective measure was used to confirm their answers. One could have used an independent measure, such as an ActiGraph (ActiGraph LLC, Pensacola, FL, USA), several days preceding the experimental session, to gain a better understanding of the sleep patterns of the subject and obtain an objective measure of the amount of sleep the subject had prior to the experimental session to confirm that the subject was not sleep-deprived. Similarly, it would have been beneficial to test subjects to verify that they had refrained from consuming stimulants or depressants during the prescribed period prior to the experimental session, rather than relying on their compliance and honesty.

The nature of the tracking task was such that there were certain segments of the target waveform (i.e., flat spots) that did not require the subject to make corrective movements to keep the cursor on the target. This meant that a measure of true visuomotor performance could not be obtained during these instances. For example, subjects could have brief behavioural microsleeps during flat spots which could not be detected via abrupt deterioration in tracking performance. A means to prevent this limitation would be to use a tracking task which has a lower likelihood of target flat spots and, hence, requires the user to always actively track the target, such as the 2-D compensatory tracking task used by Makeig and Jolley (1996).

Another limitation is that the reliability of the video rating was not determined as only one investigator (MP) rated all the video recordings, with no measure of inter-rater or intra-rater reliability. However, a member of the Lapse Research Programme informally re-evaluated a sub-group (a session each from 5 subjects) of video recordings and verified that there were no substantial discrepancies in the rating. The main reason for not carrying out inter-rater evaluations of the video recordings was simply due to the extensive additional time required. In future, it would be useful to have additional scorers rate the video using a common set of rules.

Only conservative estimates of lapses (video BMs and only flat spots in tracking) were intentionally chosen to characterize and detect lapses in this project. Instances where the target was moving but the user response was non-coherent were deliberately excluded mainly due to difficulties in defining their start and end points. However, it is important to note that non-coherent tracking is also a type of lapse in responsiveness and, the exclusion of such events has led to an underestimation of the number of lapses.

The exclusion of artifact contaminated epochs in lapse detection model testing is likely to have biased the model towards a better performance. However, as mentioned in § 8.11, the

proportion of excluded epochs was relatively low (8.5%) and hence unlikely to have substantially worsened performance if included.

On a positive note, the experimental protocol of the study was such that lapses in the form of BMs occurred in 8 out of the 15 normal non-sleep-deprived young subjects to a substantial extent during normal working hours. Factors such as a reasonably long session length (1 h), repeated sessions (×2), scheduling the sessions after lunch during the circadian low, and use of a warm, quiet, and dark room for the experiment all contributed towards achieving the goal of capturing numerous lapses. While it was hoped that some subjects would have lapses in performance, the extent to which this occurred in non-sleep-deprived subjects was quite unexpected. This study provided an extremely valuable dataset of EEG, EOG, video, and tracking performance data containing definite complete lapses in performance, including behavioural microsleeps. The value of this dataset has been well demonstrated in the quantification of characteristics of lapses (Peiris *et al.*, 2006) and EEG-based detection of lapses with high temporal resolution (Davidson *et al.*, 2007). It would have been even better to have had a larger number of subjects for the training and validation of the detection models, enabling one to arrive at more definitive conclusions and, all going well, superior detection performance.

The work documented in Chapter 7 is the first to provide detailed information about the characteristics of complete lapses of responsiveness during an extended CTT in non-sleep-deprived subjects.

Although modest, the spectral power based lapse detector (described in Chapter 8) has shown comparable performance to other more complex systems despite its relatively simple structure, using only linear features of the EEG such as spectral power.

9.5 Future work

The tracking task used in the CTT study is not, and was not intended to be, a true driving simulator. Hence it is not possible to extrapolate the findings of that study in terms of the number of lapses to a real-life driving situation. This is because a real driving task tends to be more stimulating to the subject and, hence, is more likely to keep their attention focused on the task. The most important difference between testing a subject in a simulator and testing them on-road are the *consequences* of lapsing between the two situations. This aspect of lapsing could be investigated further using an on-road driving test – e.g., Papadelis *et al*. (2007).

As mentioned in the Critique section, future lapse studies should utilize an independent measure such as an ActiGraph to obtain an objective measure of the amount of sleep a subject had in the days preceding an experimental session. Tests should also be conducted to confirm that subjects have refrained from consuming stimulants or depressants during the stipulated time period. The task used to measure subject performance should require the subject to always participate actively (i.e., no target flat spots).

A wider cross-section of the population should also be selected for any future studies of this nature. This would enable one to generalize the findings such as the characteristics of lapses to a broader population and determine differences due to age, gender, etc. Similarly, detection models formed using data from a wider demographic base will presumably generalize better across the population. It would also be pertinent to recruit a larger number of subjects to (a) increase the quantity of available data to create a more generalized detector (b) increase the amount of data available to validate the detector, and (c) determine the effects of age, sex, sleep-deprivation, food/drink, etc., on lapsing.

The use of two cameras may also prove useful whereby one camera could provide a close-up view of the subject's face while the second camera could be used to capture a more wide-angled view. The close-up camera could be used to observe aspects such as blink rate and duration as well as calculate related measures such as PERCLOS (Wierwille and Ellsworth, 1994). The wide-angled view could be used to obtain additional cues such as body posture and arm/leg position of the subject.

Despite formidable technical hurdles, the use of functional-MRI during such a study should also be given due consideration as it will shed light on brain mechanisms underlying lapses. Plans are underway to carry out such a project by a group of members of the Lapse Research Programme (Poudel *et al.*, 2006).

The practicality of using EEG in the detection of lapses in performance in real-life situations should also be re-evaluated (Papadelis *et al.*, 2007). EEG was primarily recorded in this study to determine if it contains cues which could be used to detect the onset of lapses reliably and, if possible, predict their onset by several seconds. However, no such reliable cues were identified and, as far as the author is aware, none have been reported in the literature. One must also consider the practical aspect of obtaining a reasonably clean EEG signal during real-life driving situations. Presumably, a cap fitted with dry electrodes may be used. However, the quality of the signals that can be obtained from such a device must be reasonably high to allow an EEG-based detection algorithm to function effectively. The operators may also be reluctant to wear additional gear while carrying out their job (such as driving). However, it is important to emphasize that not all lapses are evident from facial

changes (especially eye closure). Examples of such instances were found in the CTT study in at least one subject (although not formally quantified). Also, video-based detection cannot possibly detect impending lapses and can only detect lapses a second or two after microsleeps commence, which may be too late from a safety point of view.

EEG data from the CTT study could be rated by experts according to a standard set of criteria, such as R&K, or, better still, two sets of criteria, such as the R&K and the scale proposed by Hori *et al*. (2001). This information could be used to determine the correlation between behavioural alertness (according to task performance and video rating) and physiological arousal as indicated by EEG. Additionally, the tracking performance can be compared with the EEG rating to identify the proportion of various types of lapses in each arousal stage.

Future work could also further investigate video lapses which do not contain flats spots, to determine whether there is any noticeable deterioration in performance, such as erratic tracking, or if the subject adequately tracks the target despite appearing to have a behavioural microsleep. Conversely, instances where the subject appears to be alert video-wise but shows poor or no tracking performance (i.e., reflected either as flat spots or erratic tracking) need to be investigated to rule out cases of distraction. Incidents where tracking appears poor, despite the video showing that subject was alert, are of special interest as these may provide evidence on the existence of instances where the subject can experience a lapse in performance despite externally appearing to be alert and not distracted. This type of lapse may require EEG data in order to be detected, as eye movement and video-based methods will not reveal the event.

Detector models could be re-examined without excluding epochs contaminated with artifacts to obtain a more realistic estimate of detector performance. One possible approach might be to mark a detector epoch contaminated with an artifact (according to the previous z-score criteria) as a 'no-lapse' irrespective of the model output. This would ensure that the detector would not mark such contaminated epochs as false positives (even though it may occasionally miss some true positives during the process).

Continuing on from the work described in § 8.10.8, it would be interesting to determine the relationship between lapse detection accuracy and lapse duration for various detection thresholds. Avenues of improving the method used to determine the performance of a lapse event detector which has variable duration events (§ 8.10.7) could also be pursued.

It is of interest to determine how well a detection model based on eye movements alone would perform and compare its performance to the EEG-based detectors that were

constructed during this project. It is likely that an EEG and EOG feature-based detector would perform better than detectors based on either of the features alone.

The question might be asked as to why wavelet decomposition was not used to analyse the EEG data from the CTT when forming models. This approach was, in fact, considered and used during initial analysis of the ATC dataset. However, initial results showed that the wavelet approach did not provide superior features over the more simple spectral power approach. Additionally, wavelet analysis produces a large quantity of features demanding increased processor power and memory capacity, as well as increasing the complexity of detection models. Therefore, it was decided to move forward with the simpler approach. However, since analysis showed moderately good detector performance with models formed using EEG features extracted via power spectral analysis, it would be of interest to see whether the addition of wavelet features to the model could indeed improve detection performance.

Signal processing techniques such as higher-order spectra should also be investigated to see if they can make a substantial contribution to improved lapse detection. They have been shown to be of value in other EEG pattern-recognition tasks such as detecting state transitions in sleep spindles (Akgul *et al.*, 2000) and differentiating epileptic EEG from normal EEG with high confidence levels (Chua *et al.*, 2007). The use of independent components derived from the EEG signals might also improve detector performance, as suggested by Lin *et al*. (2006; 2005b), who used it to estimate changes in drowsiness.

References

Abasolo, D., Hornero, R., Gomez, C., Garcia, M. and Lopez, M. Analysis of EEG background activity in Alzheimer's disease patients with Lempel-Ziv complexity and central tendency measure. *Medical Engineering and Physics,* 2006, 28: 315-322.

Accardo, A., Affinito, M., Carrozzi, M. and Bouquet, F. Use of the fractal dimension for the analysis of electroencephalographic time series. *Biological Cybernetics,* 1997, 77: 339-350.

Achermann, P., Hartmann, R., Gunzinger, A., Guggenbuhl, W. and Borbely, A. A. Correlation dimension of the human sleep electroencephalogram: cyclic changes in the course of the night. *European Journal of Neuroscience,* 1994, 6: 497-500.

Afifi, A. A. and Clarke, V. *Computer-aided multivariate analysis,* Chapman & Hall, London, 1998.

Aftanas, L. I. and Golocheikine, S. A. Non-linear dynamic complexity of the human EEG during meditation. *Neuroscience Letters,* 2002, 330: 143-146.

Agarwal, R. and Gotman, J. Computer-assisted sleep staging. *IEEE Transactions in Biomedical Engineering,* 2001, 48: 1412-1423.

Åkerstedt, T. Shift work and disturbed sleep/wakefulness. *Occupational Medicine,* 2003, 53: 89-94.

Åkerstedt, T. and Gillberg, M. Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience,* 1990, 52: 29-37.

Åkerstedt, T., Peters, B., Anund, A. and Kecklund, G. Impaired alertness and performance driving home from the night shift: a driving simulator study. *Journal of Sleep Research,* 2005, 14: 17-20.

Akgul, T., Sun, M., Sclabassi, R. J. and Cetin, A. E. Characterization of sleep spindles using higher order statistics and spectra. *IEEE Transactions on Biomedical Engineering,* 2000, 47: 997-1009.

Angus, R. G., Pigeau, R. A. and Heslegrave, R. J. Sustained operation studies: from the field to the laboratory. In: C. Stampi (Ed) *Why we nap: evolution, chronobiology, and functions of polyphasic and ultrashort sleep.* Birkhauser, Boston, 1992: pp. 217-244.

Aoki, F., Fetz, E. E., Shupe, L., Lettich, E. and Ojemann, G. A. Increased gamma-range activity in human sensorimotor cortex during performance of visuomotor tasks. *Clinical Neurophysiology,* 1999, 110: 524-537.

Aoki, F., Fetz, E. E., Shupe, L., Lettich, E. and Ojemann, G. A. Changes in power and coherence of brain activity in human sensorimotor cortex during performance of visuomotor tasks. *Biosystems,* 2001, 63: 89-99.

Arnedt, J. T., Geddes, M. A. C. and MacLean, A. W. Comparative sensitivity of a simulated driving task to self-report, physiological, and other performance measures during prolonged wakefulness. *Journal of Psychosomatic Research,* 2005, 58: 61-71.

Arnedt, J. T., Wilde, G. J., Munt, P. W. and MacLean, A. W. How do prolonged wakefulness and alcohol compare in the decrements they produce on a simulated driving task? *Accident Analysis and Prevention,* 2001, 33: 337-344.

Arruda, J. E., Walker, K. A., Weiler, M. D. and Valentino, D. A. Validation of a right hemisphere vigilance system as measured by principal component and factor analyzed quantitative electroencephalogram. *International Journal of Psychophysiology,* 1999, 32: 119-128.

Aschoff, J. On the perception of time during prolonged temporal isolation. *Human Neurobiology,* 1985, 4: 41-52.

Aschoff, J. Human perception of short and long time intervals: its correlation with body temperature and the duration of wake time. *Journal of Biological Rhythms,* 1998, 13: 437-42.

Badia, P., Wright Jr., K. P. and Wauquier, A. Fluctuations in single-Hertz EEG activity during the transition to sleep. In: R. D. Ogilvie (Ed) *Sleep onset: Normal and abnormal processes.* American Psychological Association, Washington, DC, 1994.

Bahrami, B., Seyedsadjadi, R., Babadi, B. and Noroozian, M. Brain complexity increases in mania. *Neuroreport,* 2005, 16: 187-191.

Barbe, Pericas, J., Munoz, A., Findley, L., Anto, J. M. and Agusti, A. G. Automobile accidents in patients with sleep apnea syndrome. An epidemiological and mechanistic study. *American Journal of Respiratory and Critical Care Medicine,* 1998, 158: 18-22.

Barbone, F., McMahon, A. D., Davey, P. G., Morris, A. D., Reid, I. C., McDevitt, D. G. and MacDonald, T. M. Association of road-traffic accidents with benzodiazepine use. *Lancet,* 1998, 352: 1331-1336.

Barger, L. K., Cade, B. E., Ayas, N. T., Cronin, J. W., Rosner, B., Speizer, F. E. and Czeisler, C. A. Extended work shifts and the risk of motor vehicle crashes among interns. *New England Journal of Medicine,* 2005, 352: 125-134.

Begg, R. *Recurrent neural networks: application to lapse detection*. Department of Mathematics and Statistics, University of Canterbury, Christchurch, 2003.

Belyavin, A. and Wright, N. A. Changes in electrical activity of the brain with vigilance. *Electroencephalography and Clinical Neurophysiology,* 1987, 66: 137-144.

Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R. and Lopez, M. E. Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems*, 2006, 7: 63-77.

Bills, A. G. Blocking: a new principle of mental fatigue. *American Journal of Psychology,* 1931, 43: 230-245.

Bishop, C. M. *Neural Networks for Pattern Recognition,* Oxford University Press, 1995.

Bittner, R., Smrcka, P., Vysoký, P., Haná, K., Pousek, L. and Scheib, P. *Detecting of fatigue states of a car driver,* Springer, Berlin/Heidelberg, 2000.

Bonnet, M. H. Sleep deprivation. In: M. Kryger, T. Roth and W. C. Dement (Eds) *Principles and practice of sleep medicine.* Saunders, Philadelphia, 1994.

Bonnet, M. H. and Arand, D. L. We are chronically sleep deprived. *Sleep,* 1995, 18: 908-911.

Brice, C. F. and Smith, A. P. Effects of caffeine on mood and performance: a study of realistic consumption. *Psychopharmacology,* 2002, 164: 188-192.

Brieman, L. Bagging predictors. *Machine Learning,* 1996a, 24: 123-140.

Brieman, L. Stacked regressions. *Machine Learning,* 1996b, 24: 49-64.

Bronzino, J. D., Berbari, E. J., Johnson, P. L. and Smith, W. M. Bioelectronics and instruments. In: R. C. Dorf (Ed) *The electrical engineering handbook.* CRC Press, 2000.

Brown, I. D. Prospects for technological countermeasures against driver fatigue. *Accident Analysis and Prevention,* 1997, 29: 525-531.

Bruhn, J., Ropcke, H. and Hoeft, A. Approximate entropy as an electroencephalographic measure of anesthetic drug effect during desflurane anesthesia. *Anesthesiology,* 2000, 92: 715-726.

Bullmore, E. T., Brammer, M. J., Bourlon, P., Alarcon, G., Polkey, C. E., Elwes, R. and Binnie, C. D. Fractal analysis of electroencephalographic signals intracerebrally recorded during 35 epileptic seizures: evaluation of a new method for synoptic visualisation of ictal events. *Electroencephalography and Clinical Neurophysiology,* 1994, 91: 337-345.

Burioka, N., Cornelissen, G., Halberg, F., Kaplan, D. T., Suyama, H., Sako, T. and Shimizu, E. Approximate entropy of human respiratory movement during eye-closed waking and different sleep stages. *Chest,* 2003, 123: 80-86.

Burioka, N., Miyata, M., Cornelissen, G., Halberg, F., Takeshima, T., Kaplan, D. T., Suyama, H., Endo, M., Maegaki, Y., Nomura, T., Tomita, Y., Nakashima, K. and Shimizu, E. Approximate entropy in the electroencephalogram during wake and sleep. *Clinical EEG and Neuroscience,* 2005, 36: 21-24.

Caffarel, J., Gibson, G. J., Harrison, J. P., Griffiths, C. J. and Drinnan, M. J. Comparison of manual sleep staging with automated neural network-based analysis in clinical practice. *Medical and Biological Engineering and Computing,* 2006, 44: 105-110.

Cajochen, C., Brunner, D. P., Krauchi, K., Graw, P. and Wirz-Justice, A. Power density in theta/alpha frequencies of the waking EEG progressively increases during sustained wakefulness. *Sleep,* 1995, 18: 890-894.

Cajochen, C., Khalsa, S. B., Wyatt, J. K., Czeisler, C. A. and Dijk, D. J. EEG and ocular correlates of circadian melatonin phase and human performance decrements during sleep loss. *American Journal of Physiology,* 1999, 277: 640-649.

Caldwell, J. A. The impact of fatigue in air medical and other types of operations: A review of fatigue facts and potential countermeasures. *Air Medical Journal,* 2001, 20: 25-32.

Caldwell, J. A. Fatigue in aviation. *Travel Medicine and Infectious Disease,* 2005, 3: 85-96.

Carlson, N. R. *Physiology of behaviour,* Pearson Education Inc., Boston, 2004.

Carskadon, M. A. and Dement, W. C. Normal human sleep: an overview. In: M. Kryger, T. Roth and W. C. Dement (Eds) *Principles and practice of sleep medicine.* W. B. Saunders Company, Philadelphia, 2000: pp. 15-25.

Carskadon, M. A., Dement, W. C., Mitler, M. M., Roth, T., Westbrook, P. R. and Keenan, S. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. *Sleep,* 1986, 9: 519-524.

Carskadon, M. A. and Rechtschaffen, A. Monitoring and staging human sleep: methodology. In: M. Kryger, T. Roth and W. C. Dement (Eds) *Principles and practice of sleep medicine.* W. B. Saunders Company, Philadelphia, 2000: pp. 15-25.

Chua, C. K., Chandran, V., Acharya, R. and Lim, C. M. Higher Order Spectral (HOS) Analysis Of Epileptic EEG Signals in *IEEE Engineering in Medicine and Biology Society*, 2007. 6495-6498.

Cluydts, R., De Valck, E., Verstraeten, E. and Theys, P. Daytime sleepiness and its evaluation. *Sleep Medicine Reviews,* 2002, 6: 83-96.

Cohen, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement,* 1960, 20: 37-46.

Connor, J., Norton, R., Ameratunga, S., Robinson, E., Civil, I., Dunn, R., Bailey, J. and Jackson, R. Driver sleepiness and risk of serious injury to car occupants: population based case control study. *British Medical Journal,* 2002, 324: 1125.

Cummings, P., Koepsell, T. D., Moffat, J. M. and Rivara, F. P. Drowsiness, counter-measures to drowsiness, and the risk of a motor vehicle crash. *Injury Prevention,* 2001, 7: 194-199.

Curcio, G., Casagrande, M. and Bertini, M. Sleepiness: evaluating and quantifying methods. *International Journal of Psychophysiology,* 2001, 41: 251-263.

Danker-Hopfe, H., Kunz, D., Gruber, G., Klosch, G., Lorenzo, J. L., Himanen, S. L., Kemp, B., Penzel, T., Roschke, J., Dorn, H., Schlogl, A., Trenker, E. and Dorffner, G. Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders. *Journal of Sleep Research,* 2004, 13: 63-69.

Davidson, P. R., Jones, R. D. and Peiris, M. T. R. EEG-based Lapse Detection with High Temporal Resolution. *IEEE Transactions in Biomedical Engineering,* 2007, 54: 832-839.

Davies, D. R. and Parasuraman, R. *The psychology of vigilance,* Academic Press, London, 1982.

Davis, H., Davis, P. A., Loomis, A. L., Harvey, E. N. and Hobart, G. Changes in human brain potentials during the onset of sleep. *Science,* 1937, 86: 448-450.

Davis, H., Davis, P. A., Loomis, A. L., Harvey, E. N. and Hobart, G. Human brain potentials during the onset of sleep. *Journal of Neurophysiology,* 1938, 1: 24-38.

Dawson, D. and Reid, K. Fatigue, alcohol and performance impairment. *Nature,* 1997, 388: 235.

De Gennaro, L., Ferrara, M., Curcio, G. and Cristiani, R. Antero-posterior EEG changes during the wakefulness–sleep transition. *Clinical Neurophysiology,* 2001, 112: 1901-1911.

De Gennaro, L., Ferrara, M., Ferlazzo, F. and Bertini, M. Slow eye movements and EEG power spectra during wake-sleep transition. *Clinical Neurophysiology,* 2000, 111: 2107-2115.

De Gennaro, L., Ferrara, M., Vecchio, F., Curcio, G. and Bertini, M. An electroencephalographic fingerprint of human sleep. *Neuroimage,* 2005, 26: 114-122.

Delorme, A. and Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods,* 2004, 134: 9-21.

Dement, W. C. and Kleitman, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology,* 1957, 9: 673-690.

Dinges, D. F. and Kribbs, N. B. Performing while sleepy: Effects of experimentally-induced sleepiness. In: T. H. Monk (Ed) *Sleep, sleepiness and performance.* John Wiley and Sons, Chichester, 1991.

Dinges, D. F., Mallis, M. M., Maislin, G. and Powell, J. W. *Evaluation of techniques for ocular measurement as an index of fatigue and as the basis for alertness management*. Final Report DOT HS 808 762. National Highway Traffic Safety Administration, Washington, DC, 1998.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., Aptowicz, C. and Pack, A. I. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep,* 1997, 20: 267-277.

Dinges, D. F. and Powell, J. W. Microcomputer analysis of performance on a portable, simple visual RT task during sustained operation. *Behavior Research Methods, Instruments, & Computers,* 1985, 17: 652-655.

Dingle, A. A., Jones, R. D., Carroll, G. J. and Fright, W. R. A multi-stage system to detect epileptiform activity in the EEG. *IEEE Transactions on Biomedical Engineering,* 1993, 40: 1260-1268.

Doane, M. G. Interactions of eyelids and tears in corneal wetting and the dynamics of the normal human eyeblink. *American Journal of Ophthalmology,* 1980, 89: 507-516.

Doran, S. M., Van Dongen, H. P. and Dinges, D. F. Sustained attention performance during sleep deprivation: evidence of state instability. *Archives Italiennes de Biologie,* 2001, 139: 253-267.

Dorrian, J., Rogers, N. L. and Dinges, D. F. Psychomotor vigilance performance: neurocognitive assay sensitive to sleep loss. In: C. A. Kushida (Ed) *Sleep deprivation - clinical issues, pharmacology, and sleep loss effects.* Marcel/Dekker, New York, 2005.

Dove, R. *Long Term EEG Users Guide*. Medical Physics & Bioengineering, Christchurch Hospital, Christchurch, 2000.

Duffy, F. H., Iyer, V. G. and Surwillo, W. W. *Clinical electroencephalography and topographic brain mapping,* Springer-Verlag, 1989.

Elbert, T., Ray, W. J., Kowalik, Z. J., Skinner, J. E., Graf, K. E. and Birbaumer, N. Chaos and physiology: deterministic chaos in excitable cell assemblies. *Physiological Reviews,* 1994, 74: 1-47.

Elger, C. E., Widman, G., Andrzejak, R., Arnhold, J., David, P. and Lehnertz, K. Nonlinear EEG analysis and its potential role in epileptology. *Epilepsia,* 2000, 41 Suppl 3: S34-8.

Empson, J. A. C. *Human brainwaves: The psychological significance of the electroencephalogram,* Stockton Press, New York, 1986.

Everitt, B. S. *Making sense of statistics in psychology: a second-level course,* Oxford University Press, New York, 1996.

Fairclough, S. H. and Graham, R. Impairment of driving performance caused by sleep deprivation or alcohol: a comparative study. *Human Factors,* 1999, 41: 118-128.

Fell, J., Roschke, J., Mann, K. and Schaffner, C. Discrimination of sleep stages: a comparison between spectral and nonlinear EEG measures. *Electroencephalography and Clinical Neurophysiology,* 1996, 98: 401-410.

Ferrara, M. and De Gennaro, L. How much sleep do we need? *Sleep Medicine Reviews,* 2001, 5: 155-179.

Ferrario, M., Signorini, M. G., Magenes, G. and Cerutti, S. Comparison of entropy-based regularity estimators: application to the fetal heart rate signal for the identification of fetal distress. *IEEE Transactions on Biomedical Engineering,* 2006, 53: 119-125.

Ferri, R., Alicata, F., Del Gracco, S., Elia, M., Musumeci, S. A. and Stefanini, M. C. Chaotic behavior of EEG slow-wave activity during sleep. *Electroencephalography and Clinical Neurophysiology*, 1996, 99: 539-543.

Findley, L., Unverzagt, M., Guchu, R., Fabrizio, M., Buckner, J. and Suratt, P. Vigilance and automobile accidents in patients with sleep apnea or narcolepsy. *Chest,* 1995, 108: 619-624.

Fisher, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics,* 1936, 7: 179-188.

Folkard, S. Black times: Temporal determinants of transport safety. *Accident Analysis and Prevention,* 1997, 29: 417-430.

Folstein, M. F. and Luria, R. Reliability, validity, and clinical application of the Visual Analogue Mood Scale. *Psychological Medicine,* 1973, 3: 479-486.

Freund, D. M., Knipling, R. R., Landsburg, A. C., Simmons, R. R. and Thomas, G. R. *A holistic approach to operator alertness research*. 950636. U.S. Department of Transportation, Washington, D. C., 1995.

Galinsky, T. L., Rosa, R. R., Warm, J. S. and Dember, W. N. Psychophysiological determinants of stress in sustained attention. *Human Factors,* 1993, 35: 603-614.

Gillberg, M. and Åkerstedt, T. Sleep loss and performance: no "safe" duration of a monotonous task. *Physiology and Behavior,* 1998, 64: 599-604.

Glenville, M., Broughton, R., Wing, A. M. and Wilkinson, R. T. Effects of sleep deprivation on short duration performance measures compared to the Wilkinson auditory vigilance task. *Sleep*, 1978, 1: 169-176.

Grace, R. Drowsy driver monitor and warning system in *International Driving Symposium. Human Factors in Driver Assessment, Training and Vehicle Design*, 2001.

Grace, R., Byrne, V. E., Bierman, D. M., Legrand, J. M., Gricourt, D. and Davis, B. K. A drowsy driver detection system for heavy vehicles in *AIAA/IEEE/SAE Digital Avionics Systems Conference*, 1998.

Grandjean, E. Fatigue in industry. *British Journal of Internal Medicine,* 1979, 36: 175-186.

Grassberger, P. and Procaccia, I. Estimation of the Kolmogorov entropy from a chaotic signal. *Physical Review A,* 1983a, 28: 2591-2593.

Grassberger, P. and Procaccia, I. Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena,* 1983b, 9: 189-208.

Greenhouse, S. W. and Geisser, S. On methods in the analysis of profile data. *Psychometrika,* 1959, 24: 95-112.

Harrison, Y. and Horne, J. A. Occurrence of 'microsleeps' during daytime sleep onset in normal subjects. *Electroencephalography and Clinical Neurophysiology,* 1996, 98: 411-416.

Hemmeter, U., Bischof, R., Hatzinger, M., Seifritz, E. and Holsboer-Trachsler, E. Microsleep during partial sleep deprivation in depression. *Biological Psychiatry,* 1998, 43: 829-839.

Herscovitch, J. and Broughton, R. Sensitivity of the Stanford sleepiness scale to the effects of cumulative partial sleep deprivation and recovery oversleeping. *Sleep,* 1981, 4: 83-91.

Higuchi, T. Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena,* 1988, 31: 277-283.

Hockey, R. Varieties of attentional state: The effects of environment. In: R. Parasuraman and D. R. Davies (Eds) *Varieties of attention.* Academic Press, Orlando, FL, 1984: pp. 449-483.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R. and Dement, W. C. Quantification of sleepiness: a new approach. *Psychophysiology,* 1973, 10: 431-436.

Hodgkin, A. L. and Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol,* 1952, 117: 500-44.

Hori, T., Hayashi, M. and Morikawa, T. Topographic EEG changes and the hypnagogic experience. In: R. D. Ogilvie (Ed) *Sleep onset: Normal and abnormal processes.* American Psychological Association, Washington, 1994.

Hori, T., Sugita, Y., Koga, E., Shirakawa, S., Inoue, K., Uchida, S., Kuwahara, H., Kousaka, M., Kobayashi, T., Tsuji, Y., Terashima, M., Fukuda, K. and Fukuda, N. Proposed supplements and amendments to 'A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects', the Rechtschaffen & Kales (1968) standard. *Psychiatry and Clinical Neurosciences,* 2001, 55: 305-310.

Horne, J. *Why we sleep,* Oxford University Press, New York, 1988.

Horne, J. A. and Pettitt, A. N. High incentive effects on vigilance performance during 72 hours of total sleep deprivation. *Acta Psychologica,* 1985, 58: 123-139.

Horne, J. A. and Reyner, L. A. Driver sleepiness. *Journal of Sleep Research,* 1995a, 4: 23-29.

Horne, J. A. and Reyner, L. A. Sleep related vehicle accidents. *British Medical Journal,* 1995b, 310: 565-567.

Horstmann, S., Hess, C. W., Bassetti, C., Gugger, M. and Mathis, J. Sleepiness-related accidents in sleep apnea patients. *Sleep,* 2000, 23: 383-389.

Huang, R., Tsai, L. and Kuo, C. J. Selection of valid and reliable EEG features for predicting auditory and visual alertness levels. *Proceedings of the National Sciences Council,* 2001, 25: 17-25.

Itil, T. M., Shapiro, D. M., Fink, M. and Kassebaum, D. Digital computer classifications of EEG sleep stages. *Electroencephalography and Clinical Neurophysiology,* 1969, 27: 76-83.

Jelles, B., van Birgelen, J. H., Slaets, J. P. J., Hekster, R. E. M., Jonkman, E. J. and Stam, C. J. Decrease of non-linear structure in the EEG of Alzheimer patients compared to healthy controls. *Clinical Neurophysiology,* 1999, 110: 1159-1167.

Jeong, J., Kim, D. J., Chae, J. H., Kim, S. Y., Ko, H. J. and Paik, I. H. Nonlinear analysis of the EEG of schizophrenics with optimal embedding dimension. *Medical Engineering and Physics,* 1998, 20: 669-676.

Ji, Q. and Yang, X. Real-Time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance. *Real-Time Imaging,* 2002, 8: 357-377.

Johns, M. W. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep,* 1991, 14: 540-545.

Johns, M. W. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep,* 1992, 15: 376-381.

Johns, M. W., Tucker, A., Chapman, R., Crowley, K. and Michael, N. Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers. *Somnologie,* 2007, 108.

Jolliffe, I. T. *Principal component analysis,* Springer-Verlag, New York, 2002.

Jones, R. D. Measurement of sensory-motor control performance capacities. In: J. D. Bronzino (Ed) *The Biomedical Engineering Handbook.* CRC Press, Boca Raton, Florida, 2000: pp. 149:1-25.

Jones, R. D., Sharman, N. B., Watson, R. W. and Muir, S. R. A PC-based battery of tests for quantitative assessment of upper-limb sensory-motor function in brain disorders. *Proceedings of the 15th International Conference of the IEEE Engineering in Medicine and Biology Society,* 1993, 15: 1414-1415.

Jung, T., Makeig, S., Stensmo, M. and Sejnowski, T. Estimating alertness from the EEG power spectrum. *IEEE Transactions on Biomedical Engineering,* 1997, 44: 60-69.

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V. and Sejnowski, T. J. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology,* 2000a, 37: 163-178.

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E. and Sejnowski, T. J. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology,* 2000b, 111: 1745-1758.

Kahnemann, D. *Attention and effort*, Prentice Hall, Englewood Cliffs, NJ, 1973.

Kamp, A. and Lopes da Silva, F. H. Technological basis of EEG recording. In: E. Niedermeyer and F. H. Lopes da Silva (Eds) *Electroencephalography: Basic principles, clinical applications and related fields.* Williams & Wilkins, Baltimore, 1999: pp. 110-121.

Kay, G. G. and Quig, M. E. Impact of sedating antihistamines on safety and productivity. *Allergy and Asthma Proceedings,* 2001, 22: 281-283.

Kiefte, M. (1999) *Discriminant analysis toolbox*, 0.3 for MATLAB.

Kim, D.-J., Jeong, J., Chae, J.-H., Park, S., Yong Kim, S., Jin Go, H., Paik, I.-H., Kim, K.-S. and Choi, B. An estimation of the first positive Lyapunov exponent of the EEG in patients with schizophrenia. *Psychiatry Research: Neuroimaging,* 2000, 98: 177-189.

Kleitman, N. *Sleep and wakefulness,* University of Chicago Press, Chicago, 1963.

Klonowski, W., Olejarczyk, E., Stepien, R., Jalowiecki, P. and Rudner, R. (2005) *Methods of monitoring of the depth of anaesthesia*, P-372355, Warsaw.

Knipling, R. R. and Wang, J. S. Revised estimates of the US drowsy driver crash problem size based on general estimates system case reviews in *39th Annual Association for the Advancement of Automotive Medicine*, Chicago, IL, 1995. 451-466.

Kobayashi, T., Madokoro, S., Wada, Y., Misaki, K. and Nakagawa, H. Human sleep EEG analysis using the correlation dimension. *Clinical Electroencephalography,* 2001, 32: 112-118.

Koch, C. and Laurent, G. Complexity and the nervous system. *Science,* 1999, 284: 96-98.

Kojima, T., Shimasono, Y., Ichise, K., Atsumi, Y., Ando, H. and Auto, K. Eye movement as an indicator of brain function. *Folia Psychiatrica et Neurologica Japonica,* 1981, 35: 425-436.

Kooi, K. A., Guvener, A. M., Tupper, C. J. and Bagchi, B. K. Electroencephalographic patterns of the temporal region in normal adults. *Neurology,* 1964, 14: 1029-1035.

Kotsiantis, S. B. and Pintelas, P. E. Hybrid feature selection instead of ensembles of classifiers in medical decision support in *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Perugia, Italy, 2004.

Kuhlo, W. and Lehmann, D. Das einschlaferleben und seine neurophysiologischen korrelate. *Archiv fur Psychiatrie und Nervenkrankheiten,* 1964, 205: 687-716.

Kuwahara, H., Higashi, H., Mizuki, Y., Matsunari, S., Tanaka, M. and Inanaga, K. Automatic real-time analysis of human sleep stages by an interval histogram method. *Electroencephalography and Clinical Neurophysiology,* 1988, 70: 220-229.

Lagarde, D., Batejat, D., Van Beers, P., Sarafian, D. and Pradella, S. Interest of modafinil, a new psychostimulant, during a sixty-hour sleep deprivation experiment. *Fundamental and Clinical Pharmacology,* 1995, 9: 271-279.

Lal, S. K. L. and Craig, A. A critical review of the psychophysiology of driver fatigue. *Biological Psychology,* 2001, 55: 173-194.

Lal, S. K. L. and Craig, A. Driver fatigue: Electroencephalography and psychological assessment. *Psychophysiology,* 2002, 39: 313-321.

Lal, S. K. L. and Craig, A. Reproducibility of the spectral components of the electroencephalogram during driver fatigue. *International Journal of Psychophysiology,* 2005, 55: 137-143.

Lal, S. K. L., Craig, A., Boord, P., Kirkup, L. and Nguyen, H. Development of an algorithm for an EEG-based driver fatigue countermeasure. *Journal of Safety Research,* 2003, 34: 321-328.

Lamond, N. and Dawson, D. Quantifying the performance impairment associated with fatigue. *Journal of Sleep Research,* 1999, 8: 255-262.

Landis, J. R. and Koch, G. C. The measurement of observer agreement for categorical data. *Biometrics,* 1977, 33: 1089-1091.

Larsen, L. E. and Walter, D. O. On automatic methods of sleep staging by EEG spectra. *Electroencephalography and Clinical Neurophysiology,* 1970, 28: 459-467.

Le Van Quyen, M., Martinerie, J., Baulac, M. and Varela, F. Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings. *Neuroreport,* 1999, 10: 2149-2155.

Le Van Quyen, M., Martinerie, J., Navarro, V., Boon, P., D'Have, M., Adam, C., Renault, B., Varela, F. and Baulac, M. Anticipation of epileptic seizures from standard EEG recordings. *The Lancet,* 2001, 357: 183-188.

Lehnertz, K. Non-linear time series analysis of intracranial EEG recordings in patients with epilepsy — an overview. *International Journal of Psychophysiology,* 1999, 34: 45-52.

Lempel, A. and Ziv, J. On the complexity of finite sequences. *IEEE Transactions on Information Theory,* 1976, IT-22: 75-81.

Lenne, M. G., Triggs, T. J. and Redman, J. R. Interactive effects of sleep deprivation, time of day, and driving experience on a driving task. *Sleep,* 1998, 21: 38-44.

Levine, B., Roehrs, T., Zorick, F. and Roth, T. Daytime sleepiness in young adults. *Sleep,* 1988, 11: 39-46.

Liberson, W. T. and Liberson, C. W. EEG records, reaction times, eye movements, respiration, and mental content during drowsiness. *Recent advances in biological psychiatry,* 1965, 8: 295-302.

Lin, C.-T., Ko, L.-W., Chung, I.-F., Huang, T.-Y., Chen, Y.-C., Jung, T.-P. and Liang, S.-F. Adaptive EEG-based alertness estimation system by using ICA-based fuzzy neural networks. *IEEE Transactions on Circuits and Systems,* 2006, 53: 2469-2476.

Lin, C.-T., Wu, R.-C., Jung, T.-P., Liang, S.-F. and Huang, T.-Y. Estimating driving performance based on EEG spectrum analysis. *EURASIP Journal on Applied Signal Processing,* 2005a, 19: 3165-3174.

Lin, C.-T., Wu, R.-C., Liang, S.-F., Chao, W.-H., Chen, Y.-J. and Jung, T.-P. EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Transactions on Circuits and Systems,* 2005b, 52: 2726-2738.

Lindsley, D. B. Psychological phenomena and the electroencephalogram. *Electroencephalography and Clinical Neurophysiology,* 1952, 4: 443-456.

Lutzenberger, W., Preissl, H. and Pulvermuller, F. Fractal dimension of electroencephalographic time series and underlying brain processes. *Biological Cybernetics,* 1995, 73: 477-482.

Lyznicki, J. M., Doege, T. C., Davis, R. M. and Williams, M. A. Sleepiness, driving, and motor vehicle crashes. Council on Scientific Affairs, American Medical Association. *JAMA,* 1998, 279: 1908-1913.

Ma, Q., Ning, X., Wang, J. and Li, J. Sleep-stage characterization by nonlinear EEG analysis using wavelet-based multifractal formalism in *IEEE Engineering in Medicine and Biology Society*, Shanghai, China, 2005. 4526-4529.

Mackworth, J. F. Performance decrement in vigilance, threshold, and high-speed perceptual motor tasks. *Canadian Journal of Psychology,* 1964, 18: 209-223.

Mackworth, J. F. *Vigilance and habituation: a neuropsychological approach,* Penguin, Harmondsworth, 1969.

Makeig, S. and Inlow, M. Lapses in alertness: coherence of fluctuations in performance and EEG spectrum. *Electroencephalography and Clinical Neurophysiology,* 1993, 86: 23-35.

Makeig, S. and Jolley, K. *COMPTRACK: A compensatory tracking task for monitoring alertness*. Technical Document 96-3C. Naval Health Research Center, San Diego, 1996.

Makeig, S. and Jung, T. Changes in alertness are a principle component of variance in the EEG spectrum. *Neuroreport,* 1995, 7: 213-216.

Makeig, S. and Jung, T.-P. Tonic, phasic, and transient EEG correlates of auditory awareness in drowsiness. *Brain Research. Cognitive Brain Research,* 1996, 4: 15-25.

Makeig, S., Jung, T. P. and Sejnowski, T. Using feedforward neural networks to monitor alertness from changes in EEG correlation and coherence. In: M. Touretzky, M. Mozer and

M. Hasselmo (Eds) *Advances in Neural Information Processing Systems.* MIT Press, Cambridge, MA, 1996: pp. 931-937.

Makeig, S., Jung, T.-P. and Sejnowski, T. J. Awareness During Drowsiness: Dynamics and Electrophysiological Correlates. *Canadian Journal of Experimental Psychology,* 2000, 54: 266-273.

Mallis, M. M. (1999) *Evaluation of Techniques for Drowsiness Detection: Experiment on Performance-Based Validation of Fatigue-tracking Technologies,* Drexel University.

Martin, W. B., Johnson, L. C., Viglione, S. S., Naitoh, P., Joseph, R. D. and Moses, J. D. Pattern recognition of EEG-EOG as a technique for all-night sleep stage scoring. *Electroencephalography and Clinical Neurophysiology,* 1972, 32: 417-427.

Maruff, P., Falleti, M. G., Collie, A., Darby, D. and McStephen, M. Fatigue-related impairment in the speed, accuracy and variability of psychomotor performance: comparison with blood alcohol levels. *Journal of Sleep Research,* 2005, 14: 21-27.

Masa, J. F., Rubio, M. and Findley, L. J. Habitually sleepy drivers have a high frequency of automobile crashes associated with respiratory disorders during sleep. *American Journal of Respiratory and Critical Care Medicine,* 2000, 162: 1407-1412.

Matousek, M. and Petersen, I. A method for assessing alertness fluctuations from EEG spectra. *Electroencephalography and Clinical Neurophysiology,* 1983, 55: 108-113.

Maulsby, R. L., Kellaway, P., Graham, M., Frost, J. D. J., Proler, M. L., Low, M. D. and North, R. R. *The normative electroencephalographic data reference library. Final report, Contract NAS 9-1200*. National Aeronautics and Space Administration, 1968.

Maycock, G. Sleepiness and driving: the experience of U.K. car drivers. *Accident Analysis and Prevention,* 1997, 29: 453-462.

McCartt, A. T., Rohrbaugh, J. W., Hammer, M. C. and Fuller, S. Z. Factors associated with falling asleep at the wheel among long-distance truck drivers. *Accident Analysis and Prevention,* 2000, 32: 493-504.

McGwin, G., Jr., Sims, R. V., Pulley, L. and Roseman, J. M. Relations among chronic medical conditions, medications, and automobile crashes in the elderly: a population-based case-control study. *American Journal of Epidemiology,* 2000, 152: 424-431.

McLachlan, G. J. *Discriminant analysis and statistical pattern recognition,* Wiley-IEEE, 2005.

Mielke, P. W. and Berry, K. J. *Permutation methods: a distance function approach,* Springer, 2001.

Mitler, M. M., Gujavarty, K. S. and Browman, C. P. Maintenance of wakefulness test: a polysomnographic technique for evaluation treatment efficacy in patients with excessive somnolence. *Electroencephalography and Clinical Neurophysiology,* 1982, 53: 658-661.

Moore, D. S. and McCabe, G. P. *Introduction to the practice of statistics,* W. H. Freeman and Company, New York, 2003.

Mormann, F., Kreuz, T., Andrzejak, R. G., David, P., Lehnertz, K. and Elger, C. E. Epileptic seizures are preceded by a decrease in synchronization. *Epilepsy Research,* 2003, 53: 173-185.

Mueller, R. O. and Cozad, J. B. Standardized discriminant coefficients: a rejoinder. *Journal of Educational Statistics,* 1993, 18: 108-114.

Naatanen, R. Selective attention and evoked potentials in humans--a critical review. *Biological Psychology,* 1975, 2: 237-307.

Naidu, P. S. *Modern spectral analysis of time series,* CRC Press, 1996.

Naitoh, P. Minimal sleep to maintain performance: the search for sleep quantum in sustained operations. In: C. Stampi (Ed) *Why we nap: evolution, chronobiology, and functions of polyphasic and ultrashort sleep.* Birkhauser, Boston, 1992: pp. 199-216.

Natarajan, K., Acharya, R., Alias, F., Tiboleng, T. and Puthusserypady, K. Nonlinear analysis of EEG signals at different mental states. *BioMedical Engineering Online,* 2004, 3:

Neale, V. L., Dingus, T. A., Klauer, S. G., Sudweeks, J. and Goodman, M. An overview of the 100-car naturalistic driving study and findings. *Proceedings of the International Technical Conference on Enhanced Safety of Vehicles,* 2005, 19: 1-10.

NHTSA *Drowsy driving and automobile crashes*. DOT HS 808 707. National Highway Traffic Safety Administration expert panel on driver fatigue and sleepiness, 1998.

Niedermeyer, E. The normal EEG of the waking adult. In: E. Niedermeyer and F. H. Lopes da Silva (Eds) *Electroencephalography: Basic principles, clinical applications and related fields.* Williams & Wilkins, Baltimore, 1999: pp. 149-173.

Ogilvie, R. D. The process of falling asleep. *Sleep Medicine Reviews,* 2001, 5: 247-270.

Ogilvie, R. D., McDonagh, D. M., Stone, S. N. and Wilkinson, R. T. Eye movements and the detection of sleep onset. *Psychophysiology,* 1988, 25: 81-91.

Ogilvie, R. D. and Wilkinson, R. T. The detection of sleep onset: Behavioural and physiological convergence. *Psychophysiology,* 1984, 21: 510-520.

Oken, B. S., Salinsky, M. C. and Elsas, S. M. Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clinical Neurophysiology,* 2006, 117: 1885-1901.

Pack, A. I., Pack, A. M., Rodgman, E., Cucchiara, A., Dinges, D. F. and Schwab, C. W. Characteristics of crashes attributed to the driver having fallen asleep. *Accident Analysis and Prevention,* 1995, 27: 769-775.

Papadelis, C., Chen, Z., Kourtidou-Papadeli, C., Bamidis, P. D., Chouvarda, I. and Bekiaris, E. Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. *Clinical Neurophysiology,* 2007, 118: 1906-1922.

Parasuraman, R. and Davies, D. R. *Varieties of attention,* Academic Press, Orlando, 1984.

Parasuraman, R., Warm, J. S. and See, J. E. Brain systems of vigilance. In: R. Parasuraman (Ed) *The attentive brain.* MIT Press, Cambridge, MA, 1998: pp. 221-256.

Peiris, M. T. R., Jones, R. D., Davidson, P. R., Carroll, G., J. and Bones, P. J. Frequent lapses of responsiveness during an extended visuomotor tracking task in non-sleep-deprived subjects. *Journal of Sleep Research,* 2006, 15: 291-300.

Penetar, D., McCann, U., Thorne, D., Kamimori, G., Galinski, C., Sing, H., Thomas, M. and Belenky, G. Caffeine reversal of sleep deprivation effects on alertness and mood. *Psychopharmacology,* 1993, 112: 359-365.

Philip, P., Vervialle, F., Le Breton, P., Taillard, J. and Horne, J. A. Fatigue, alcohol, and serious road crashes in France: factorial study of national data. *British Medical Journal,* 2001, 322: 829-830.

Pigeau, R., Naitoh, P., Buguet, A., McCann, C., Baranski, J., Taylor, M., Thompson, M. and Mac, K. I. I. Modafinil, d-amphetamine and placebo during 64 hours of sustained mental work. I. Effects on mood, fatigue, cognitive performance and body temperature. *Journal of Sleep Research,* 1995, 4: 212-228.

Pincus, S. Approximate entropy (ApEn) as a complexity measure. *Chaos,* 1995, 5: 110-117.

Pincus, S. M., Gladstone, I. M. and Ehrenkranz, R. A. A regularity statistic for medical data analysis. *Journal of Clinical Monitoring,* 1991, 7: 335-345.

Pincus, S. M. and Goldberger, A. L. Physiological time-series analysis: what does regularity quantify? *American Journal of Physiology,* 1994, 266: H1643-H1656.

Pivik, R. T. The several qualities of sleepiness: Psychophysiological considerations. In: T. H. Monk (Ed) *Sleep, sleepiness and performance.* John Wiley & Sons Ltd., Chichester, 1991.

Poudel, G. R., Jones, R. D. and Davidson, P. R. Proposed multimodal study of lapses in responsiveness with simultaneous haemodynamic, electrophysiological, and behavioural recording in *The fMRI Experience*, Melbourne, 2006.

Powell, J. W. *PVT-192 and analysis software reference manual*. Unit for Experimental Psychiatry, University of Pennsylvania, Philadelphia, 1999.

Powell, N. B., Schechtman, K. B., Riley, R. W., Li, K. and Guilleminault, C. Sleepy driving: Accidents and injury. *Otolaryngology - Head and Neck Surgery,* 2002, 126: 217-227.

Powell, N. B., Schechtman, K. B., Riley, R. W., Li, K., Troell, R. and Guilleminault, C. The road to danger: the comparative risks of driving while sleepy. *Laryngoscope,* 2001, 111: 887-893.

Priest, B., Brichard, C., Aubert, G., Liistro, G. and Rodenstein, D. O. Microsleep during a simplified maintenance of wakefulness test. *American Journal of Respiratory and Critical Care Medicine,* 2001, 163: 1619-1625.

Radhakrishnan, N. and Gangadhar, B. N. Estimating regularity in epileptic seizure time-series data. A complexity measure approach. *IEEE Engineering in Medicine and Biology,* 1998, 1998: 89-94.

Ray, S. R., Lee, W. D., Morgan, C. D. and Airth-Kindree, W. Computer sleep stage scoring - an expert system approach. *International Journal of Bio-Medical Computing*, 1986, 19: 43-61.

Rechtschaffen, A. Sleep onset: Conceptual issues. In: R. D. Ogilvie (Ed) *Sleep onset: Normal and abnormal processes.* American Psychological Association, Washington, DC, 1994.

Rechtschaffen, A. and Kales, A. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. University of California, Brain Information Service/Brain Research Institute, Los Angeles, 1968.

Reyner, L. A. and Horne, J. A. Evaluation "in-car" countermeasures to sleepiness: cold air and radio. *Sleep,* 1998, 21: 46-50.

Reyner, L. A. and Horne, J. A. Early morning driver sleepiness: effectiveness of 200 mg caffeine. *Psychophysiology,* 2000, 37: 251-256.

Ripley, B. D. *Pattern recognition and neural networks*, Cambridge University Press, Cambridge, 1996.

Roge, J., Pebayle, T., El Hannachi, S. and Muzet, A. Effect of sleep deprivation and driving duration on the useful visual field in younger and older subjects during simulator driving. *Vision Research,* 2003, 43: 1465-1472.

Rosekind, M. R. Underestimating the societal costs of impaired alertness: safety, health and productivity risks. *Sleep Medicine,* 2005, 6: S21-S25.

Rupp, T., Arnedt, J. T., Acebo, C. and Carskadon, M. A. Performance on a dual driving simulation and subtraction task following sleep restriction. *Perceptual and Motor Skills,* 2004, 99: 739-753.

Sagberg, F. Road accidents caused by drivers falling asleep. *Accident Analysis and Prevention,* 1999, 31: 639-649.

Santamaria, J. and Chiappa, K. H. The EEG of Drowsiness in Normal Adults. *Journal of Clinical Neurophysiology,* 1987, 4: 327-382.

Schaltenbrand, N. Sleep stage scoring using the neural network model: comparison between visual and automatic analysis in normal subjects and patients. *Sleep*, 1996, 19: 26-35.

Schapire, R. E., Freund, Y., Bartlett, P. and Lee, W. S. Boosting the margin: a new explanation for the effectiveness of voting methods in *14th International Conference on Machine Learning*, 1997. 322-330.

Scher, M. S. Automated EEG-sleep analyses and neonatal neurointensive care. *Sleep Medicine,* 2004, 5: 533-540.

Schroeder, D. J., Touchstone, R. M., Stern, J. A., Stoliarov, N. and Thackray, R. *Maintaining vigilance on a simulated ATC monitoring task across repeated sessions*. DOT/FAA/AM-94/6. U.S. Department of Transportation - Federal Aviation Administration, 1994.

Segalowitz, S. J., Velikonja, D. and Storrie-Baker, J. Attentional allocation and capacity in waking arousal. In: R. D. Ogilvie (Ed) *Sleep onset: Normal and abnormal processes.* American Psychological Association, Washington, 1994.

Sejnowski, T. J. and Destexhe, A. Why do we sleep? *Brain Research,* 2000, 886: 208-223.

Shaw, F.-Z., Chen, R.-F., Tsao, H.-W. and Yen, C.-T. Algorithmic complexity as an index of cortical function in awake and pentobarbital-anesthetized rats. *Journal of Neuroscience Methods,* 1999, 93: 101-110.

Shen, J., Barbera, J. and Shapiro, C. M. Distinguishing sleepiness and fatigue: focus on definition and measurement. *Sleep Medicine Reviews,* 2006, 10: 63-76.

Sheskin, D. *Handbook of Parametric and Nonparametric Statistical Procedures,* CRC Press, 1997.

Shin, D. G., Yoo, C. S., Yi, S. H., Bae, J. H., Kim, Y. J., Park, J. S. and Hong, G. R. Prediction of paroxysmal atrial fibrillation using nonlinear analysis of the R-R interval dynamics before the spontaneous onset of atrial fibrillation. *Circulation Journal,* 2006, 70: 94-99.

Signal, T. L. (2002) *Scheduled napping on the night shift: consequences for the performance and neurophysiological alertness of air traffic controllers,* Sleep/Wake Research Center, University of Otago.

Signal, T. L. and Gander, P. H. Rapid counterclockwise shift rotation in air traffic control: effects on sleep and night work. *Aviation, Space, and Environmental Medicine,* 2007, 78: 878-885.

Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T. Rocr: visualizing classifier performance in r. *Bioinformatics,* 2005, 21: 3940-3941.

Skipper, J. M. and Wierwille, W. W. Drowsy driver detection using discriminant analysis. *Human Factors,* 1986, 28: 527-540.

Smith, J. D. and Karacan, I. EEG sleep stage scoring by an automatic hybrid system. *Electroencephalography and Clinical Neurophysiology,* 1971, 31: 231-237.

Sommer, D., Hink, T. and Golz, M. Application of learning vector quantization to detect drivers dozing-off in *European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems*, Albufeira, Portugal, 2002. 99-103.

Stam, C. J., Jelles, B., Achtereekte, H. A., Rombouts, S. A., Slaets, J. P. and Keunen, R. W. Investigation of EEG non-linearity in dementia and Parkinson's disease. *Electroencephalography and Clinical Neurophysiology,* 1995, 95: 309-317.

Stanus, E., Lacroix, M., Kerkhofs, M. and Mendlewicz, J. Automated sleep scoring: a comparative reliability study of algorithms. *Electroencephalography and Clinical Neurophysiology,* 1987, 66: 448-456.

Stephan, K., Hosking, S., Regan, M., Verdoon, A., Young, K. and Haworth, N. *The relationship between driving performance and the Johns drowsiness scale as measured by the Optalert system*. #0.252. Monash University Accident Research Centre, Monash University, Melbourne, 2006.

Steriade, M., McCormick, D. A. and Sejnowski, T. J. Thalamocortical oscillations in the sleeping and aroused brain. *Science,* 1993, 262: 679-685.

Stroh, C. M. *Vigilance: The problem of sustained attention*, 1971.

Stutts, J. C., Wilkins, J. W., Scott Osberg, J. and Vaughn, B. V. Driver risk factors for sleep-related crashes. *Accident Analysis and Prevention,* 2003, 35: 321-331.

Subasi, A. Application of classical and model-based spectral methods to describe the state of alertness in EEG. *Journal of Medical Systems,* 2005a, 29: 473-486.

Subasi, A. Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. *Expert Systems with Applications,* 2005b, 28: 701-711.

Takahashi, M. The role of prescribed napping in sleep medicine. *Sleep Medicine Reviews,* 2003, 7: 227-235.

Tecce, J. J., Savignano-Bowman, J. and Meinbresse, D. Contingent negative variation and the distraction--arousal hypothesis. *Electroencephalography and Clinical Neurophysiology,* 1976, 41: 277-286.

Thayer, R. E. Factor analytic and reliability studies on the Activation-Deactivation Adjective Check List. *Psychological Reports,* 1978, 42: 747-756.

Thiffault, P. and Bergeron, J. Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis and Prevention,* 2003, 35: 381-391.

Thomas, J. R., Nelson, J. K. and Silverman, S. J. *Research Methods in Physical Activity*, Human Kinetics Europe Ltd., 2005.

Thorpy, M. J. The clinical use of the Multiple Sleep Latency Test. The Standards of Practice Committee of the American Sleep Disorders Association. *Sleep,* 1992, 15: 268-76.

Tian, J. Y. and Liu, J. Q. Automated Sleep Staging by a Hybrid System Comprising Neural Network and Fuzzy Rule-based Reasoning. *Conference Proceedings of the IEEE Engineering in Medicine and Biology Society,* 2005, 4: 4115-4118.

Ting, K. M. and Witten, I. H. Issues in stacked generalization. *Journal of Artificial Intelligence Research,* 1999, 10: 271-289.

Tirunahari, V. L., Zaidi, S. A., Sharma, R., Skurnick, J. and Ashtyani, H. Microsleep and sleepiness: a comparison of multiple sleep latency test and scoring of microsleep as a diagnostic test for excessive daytime sleepiness. *Sleep Medicine,* 2003, 4: 63-67.

Tononi, G., Sporns, O. and Edelman, G. M. A measure for brain complexity: relating functional segregation and integration in the nervous system in *Proceedings of the National Academy of Sciences of the United States of America*, 1994. 5033-5037.

Torsvall, L. and Åkerstedt, T. Sleepiness on the job: continuously measured EEG changes in train drivers. *Electroencephalography and Clinical Neurophysiology,* 1987, 66: 502-511.

Torsvall, L. and Åkerstedt, T. Extreme sleepiness: quantification of EOG and spectral EEG parameters. *International Journal of Neuroscience,* 1988, 38: 435-441.

Valley, V. and Broughton, R. The physiological (EEG) nature of drowsiness and its relation to performance deficits in narcoleptics. *Electroencephalography and Clinical Neurophysiology,* 1983, 55: 243-251.

Van Orden, K. F., Jung, T.-P. and Makeig, S. Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biological Psychology,* 2000, 52: 221 - 240.

Verwey, W. B. and Zaidel, D. M. Preventing drowsiness accidents by an alertness maintenance device. *Accident Analysis and Prevention,* 1999, 31: 199-211.

Vuckovic, A., Radivojevic, V., Chen, A. C. and Popovic, D. Automatic recognition of alertness and drowsiness from EEG by an artificial neural network. *Medical Engineering and Physics,* 2002, 24: 349-360.

Walsh, J. K., Muehlbach, M. J. and Schweitzer, P. K. Simulated assembly line performance following ingestion of cetirizine or hydroxyzine. *Annals of Allergy,* 1992, 69: 195-200.

Wierwille, W. W. and Ellsworth, L. A. Evaluation of driver drowsiness by trained raters. *Accident Analysis and Prevention,* 1994, 26: 571-581.

Wilkinson, R. T. Methods for research on sleep deprivation and sleep function. *International Psychiatry Clinics,* 1970, 7: 369-381.

Wilkinson, R. T. and Houghton, D. Portable four-choice reaction time test with magnetic tape memory. *Behavior Research Methods, Instruments, and Computers,* 1975, 7: 441-446.

Wilkinson, R. T. and Houghton, D. Field test of arousal: A portable reaction timer with data storage. *Human Factors,* 1982, 24: 487-493.

Williamson, A. M., Feyer, A.-M., Mattick, R. P., Friswell, R. and Finlay-Brown, S. Developing measures of fatigue using an alcohol comparison to validate the effects of fatigue on performance. *Accident Analysis and Prevention,* 2001, 33: 313-326.

Witten, I. H. and Frank, E. *Data mining - practical machine learning tools and techniques with Java implementations,* Academic Press, 2000.

Wolpert, D. Stacked generalization. *Neural Networks,* 1992, 5: 241-259.

Wright, K., Badia, P. and Wauquier, A. EEG analysis and transition from wake: topographical and temporal patterns of brain activity during the transition from wakefulness to sleep. *Sleep,* 1995, 18: 880-889.

Yaylali, I., Kocak, H. and Jayakar, P. Detection of seizures from small samples using nonlinear dynamic system theory. *IEEE Transactions on Biomedical Engineering,* 1996, 43: 743-751.

Zhang, X. and Roy, R. J. Derived fuzzy knowledge model for estimating the depth of anesthesia. *IEEE Transactions on Biomedical Engineering,* 2001, 48: 312-323.

Zhang, X. S. and Roy, R. J. Predicting movement during anaesthesia by complexity analysis of electroencephalograms. *Medical and Biological Engineering and Computing,* 1999, 37: 327-334.

Zhang, X. S., Zhu, Y. S., Thakor, N. V. and Wang, Z. Z. Detecting ventricular tachycardia and fibrillation by complexity measure. *IEEE Transactions in Biomedical Engineering,* 1999, 46: 548-555.