CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Artefact detection and removal algorithms for EEG diagnostic systems
Authors	O'Regan, Simon H.
Publication date	2013
Original Citation	O'Regan, S. H. 2013. Artefact detection and removal algorithms for EEG diagnostic systems. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2013, Simon H. O'Regan - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2026-11-01 03:22:43
Item downloaded from	https://hdl.handle.net/10468/1391



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Artefact Detection and Removal Algorithms for EEG Diagnostic Systems

Simon O'Regan

November 17, 2013



A Thesis Submitted to the
National University of Ireland
in Fulfilment of the Requirements for
the Degree of Ph.D.

Supervisors:	Dr. William Marnane
	Dr. Gordon Lightbody
	Dr. Geraldine Boylan
Head of Department:	Dr. Nabeel Riza

Department of Electrical and Electronic Engineering,
National University of Ireland, Cork.

Abstract

The electroencephalogram (EEG) is a medical technology that is used in the monitoring of the brain and in the diagnosis of many neurological illnesses. Although coarse in its precision, the EEG is a non-invasive tool that requires minimal set-up times, and is suitably unobtrusive and mobile to allow continuous monitoring of the patient, either in clinical or domestic environments. Consequently, the EEG is the current tool-of-choice with which to continuously monitor the brain where temporal resolution, ease-of-use and mobility are important.

Traditionally, EEG data are examined by a trained clinician who identifies neurological events of interest. However, recent advances in signal processing and machine learning techniques have allowed the automated detection of neurological events for many medical applications. In doing so, the burden of work on the clinician has been significantly reduced, improving the response time to illness, and allowing the relevant medical treatment to be administered within minutes rather than hours.

However, as typical EEG signals are of the order of microvolts (μV), contamination by signals arising from sources other than the brain is frequent. These extra-cerebral sources, known as artefacts, can significantly distort the EEG signal, making its interpretation difficult, and can dramatically disimprove automatic neurological event detection classification performance.

This thesis therefore, contributes to the further improvement of automated neurological event detection systems, by identifying some of the major obstacles in deploying these EEG systems in ambulatory and clinical environments so that the EEG technologies can emerge from the laboratory towards real-world settings, where they can have a real-impact on the lives of patients. In this context, the thesis tackles three major problems in EEG monitoring, namely: (i) the problem of head-movement artefacts in ambulatory EEG, (ii) the high numbers of false detections in state-of-the-art, automated, epileptiform activity detection systems and (iii) false detections in state-of-the-art, automated neonatal seizure detection systems. To accomplish this, the thesis employs a wide range of statistical, signal processing and machine learning techniques drawn from mathematics, engineering and computer science.

The first body of work outlined in this thesis proposes a system to automatically detect head-movement artefacts in ambulatory EEG and utilises supervised machine learning

classifiers to do so. The resulting head-movement artefact detection system is the first of its kind and offers accurate detection of head-movement artefacts in ambulatory EEG. Subsequently, additional physiological signals, in the form of gyroscopes, are used to detect head-movements and in doing so, bring additional information to the head-movement artefact detection task. A framework for combining EEG and gyroscope signals is then developed, offering improved head-movement artefact detection.

The artefact detection methods developed for ambulatory EEG are subsequently adapted for use in an automated epileptiform activity detection system. Information from support vector machines classifiers used to detect epileptiform activity is fused with information from artefact-specific detection classifiers in order to significantly reduce the number of false detections in the epileptiform activity detection system. By this means, epileptiform activity detection which compares favourably with other state-of-the-art systems is achieved.

Finally, the problem of false detections in automated neonatal seizure detection is approached in an alternative manner; blind source separation techniques, complimented with information from additional physiological signals are used to remove respiration artefact from the EEG. In utilising these methods, some encouraging advances have been made in detecting and removing respiration artefacts from the neonatal EEG, and in doing so, the performance of the underlying diagnostic technology is improved, bringing its deployment in the real-world, clinical domain one step closer.

Contents

1 Introduction **1**

1.1 Introduction 1

1.2 Aim and scope of this thesis 2

1.3 Outline of thesis 3

1.4 Publications arising from thesis 5

2 EEG Artefacts **7**

2.1 Introduction 7

2.2 The electroencephalogram 7

2.3 What are EEG artefacts? 8

2.3.1 Physiological artefacts 10

2.3.2 Non-physiological artefacts 17

2.4 Artefact avoidance and minimisation 19

2.5 Biomedical signal processing and machine learning 20

2.5.1 Machine learning 20

2.5.2 Blind source separation 22

2.6 Manual and semi-automatic artefact removal 25

2.7 Automated EEG artefact detection and removal: A review 27
2.7.1 Artefact detection 27
2.7.2 Artefact removal 28
2.8 EEG artefacts in automated EEG diagnostic systems 33
2.8.1 Epileptiform activity and epileptic seizure detection 34
2.8.2 EEG artefacts in epileptiform activity detection systems 37
2.8.3 Neonatal seizure detection 39
2.8.4 EEG artefacts in neonatal seizure detection 41
2.9 Datasets and metrics 42
2.9.1 Head-movement artefacts in adult EEG 43
2.9.2 Ocular and movement artefacts in epileptiform EEG 46
2.9.3 Neonatal EEG 47
2.9.4 Performance assessment 48
2.10 Summary 52

3 Detecting artefacts arising from head-movements 54

3.1 Introduction 54
3.2 The problem of head-movement artefacts in ambulatory EEG 55
3.2.1 Generalised head-movement artefact detection: is it possible? 55
3.2.2 Ambulatory REACT system 56
3.2.3 Artefact detection using supervised machine learning 58
3.3 EEG features 59
3.3.1 Time domain features 60

3.3.2 Frequency domain features 70
3.3.3 Entropy-based features 73
3.4 Feature reduction and linear discriminant classification 75
3.4.1 Feature reduction using mutual information evaluation function . . . 76
3.4.2 MIEF results 78
3.4.3 Linear discriminant classification 79
3.4.4 LDC performance 82
3.5 Support vector machines 83
3.5.1 SVM theoretical background 84
3.5.2 SVM model selection 87
3.5.3 SVM performance 87
3.6 Summary 88

4 Detecting head-movement artefacts with gyroscopes 90

4.1 Introduction 90
4.2 Gyroscopes 90
4.3 Gyroscope features 96
4.4 Feature reduction and linear discriminant classification 103
4.4.1 Mutual information evaluation function 104
4.4.2 Linear discriminant classification 105
4.5 Support vector machines 106
4.5.1 Gyroscope SVM 108
4.5.2 Comparison of EEG and gyroscope SVM performance 109

4.6 Summary . . . 111

5 Multimodal detection of EEG head-movement artefacts 112

5.1 Introduction . . . 112

5.2 Multimodal fusion . . . 113

5.3 Feature fusion . . . 114

5.3.1 Feature fusion performance . . . 115

5.4 Score fusion . . . 115

5.4.1 Mean/Sum rule . . . 116

5.4.2 Product rule . . . 117

5.4.3 Maximum rule . . . 117

5.4.4 Minimum rule . . . 118

5.4.5 Score fusion performance . . . 118

5.5 Decision fusion . . . 119

5.5.1 Logical AND method . . . 121

5.5.2 Logical OR method . . . 121

5.5.3 Decision fusion performance . . . 122

5.6 Comparison of fusion methods: early integration vs. late integration . . . 123

5.7 Summary . . . 125

6 Robust epileptiform activity detection 126

6.1 Introduction . . . 126

6.2 Automated epilepsy detection . . . 127

6.3 Ocular and movement artefacts . . . 127
6.4 Reducing false positives without removing valuable EEG information . . . 131
6.5 Epileptiform detection classifier . . . 133
6.5.1 Detection system . . . 133
6.5.2 Preprocessing . . . 134
6.5.3 EEG feature generation . . . 134
6.5.4 Classifiers . . . 137
6.5.5 Threshold choice . . . 140
6.5.6 Metrics . . . 140
6.6 Classifier fusion architectures . . . 141
6.6.1 Score fusion . . . 142
6.6.2 Decision fusion . . . 143
6.7 Comparison of fusion methods . . . 145
6.7.1 Epoch-based metrics . . . 145
6.7.2 Event-based metrics . . . 146
6.8 Summary . . . 148

7 Artefact removal for automated neonatal seizure detection 149

7.1 Introduction . . . 149
7.2 Neonatal seizure detection . . . 150
7.2.1 Neonatal seizure . . . 150
7.2.2 Automated neonatal seizure detection . . . 150
7.3 Respiratory artefact and false detections . . . 151

7.4 Automated artefact removal . . . 153

7.5 Data . . . 156

7.6 De Vos method: correlating SOBI with respiratory trace . . . 157

7.6.1 Second order blind identification . . . 157

7.6.2 Removing respiratory artefact components . . . 159

7.6.3 Algorithm performance . . . 161

7.7 Respiratory artefact removal without a respiratory trace . . . 163

7.7.1 Electrocardiogram-derived respiration signal . . . 164

7.7.2 Respiratory artefact removal with the EDR signal . . . 166

7.7.3 Artefacts on additional physiological signals . . . 169

7.8 Summary . . . 171

8 Conclusions and future work **174**

8.1 Conclusions . . . 174

8.2 Future research directions . . . 177

8.2.1 EEG head-movement artefact detection . . . 178

8.2.2 Multimodal head-movement artefact detection . . . 178

8.2.3 Artefact processing in automated epileptiform activity detection systems . . . 179

8.2.4 Respiratory artefact removal in automated neonatal seizure detection systems . . . 181

8.3 Final remarks . . . 181

Acknowledgements **183**

Statement of originality **185**

List of Tables **186**

List of Figures **187**

References **197**

Acronyms

ANSeR	Algorithm for Neonatal Seizure Recognition
AR	Autoregressive
BCI	Brain Computer Interface
BSS	Blind Source Separation
CCA	Canonical Correlation Analysis
CTR	Conventional Trial Rejection
CUH	Cork University Hospital
CUMH	Cork University Maternity Hospital
DC	Direct Current
DFT	Discrete Fourier Transform
DSP	Digital Signal Processor
ECG	Electrocardiogram
EDR	Electrocardiogram-derived Respiration
EEDSP	Efficient Embedded Digital Signal Processing
EEG	Electroencephalogram
EMG	Electromyogram
EOG	Electrooculogram
EP	Evoked Potential
ERP	Event-related Potential
FD	False Detection
FDR	False Detection Rate
FD/h	False Detections per hour
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FN	False Negative
FP	False Positive

FT	Fourier Transform
GDR	Good Detection Rate
HIE	Hypoxic Ischemic Encephalopathy
IC	Independent Component
ICA	Independent Component Analysis
ICU	Intensive Care Unit
IMA	Infusion Motor Artefact
IWBW	Intensity Weighted Mean Bandwidth
IWMF	Intensity Weighted Mean Frequency
LD	Linear Discriminant
LDC	Linear Discriminant Classifier
LOO-CV	Leave One Out Cross Validation
MA	Moving Average
MI	Mutual Information
MIEF	Mutual Information Evaluation Function
NICU	Neonatal Intensive Care Unit
PCA	Principal Component Analysis
PDF	Probability Density Function
PSD	Power Spectral Density
REACT	Real-time EEG Analysis for Event Detection
RBF	Radial Basis Function
RMS	Root Mean Squared
RSA	Respiratory Sinus Arrhythmia
ROC	Receiver Operating Characteristic
SOBI	Second Order Blind Identification
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
UCC	University College Cork
UCLH	University College London Hospitals

Chapter 1

Introduction

1.1 Introduction

Electroencephalography (EEG) is a medical technology that is used in the monitoring of the brain and diagnosis of many neurological illnesses. By measuring voltage on or just below the scalp, EEG is the technology of choice in epilepsy and neonatal seizure detection as well as in other diagnostics such as sleep analysis. Similarly, in evoked/ event-related potentials the EEG is used to evaluate brain function, often in patients with cognitive diseases (Sanei and Chambers, 2007). In addition, many brain-computer interface (BCI) applications utilize EEG as a direct communication pathway between the brain and an external device, most commonly for assisting, augmenting, or repairing human cognitive or sensory-motor functions (Dornhege et al., 2007).

To utilize the EEG for any of the aforementioned applications requires interpretation and processing of vast quantities of information. Traditionally, EEG data is examined by a trained clinician who identifies neurological events of interest. However, recent advances in signal processing and machine learning techniques have allowed the automated detection of neurological events for many medical applications. By automating the detection of neurologically relevant events, the burden of work on the clinician can be significantly reduced, improving the response time to the illness, and allowing suitable medical treatment to be administered within minutes rather than hours (Thomas, 2011). In the case of BCI, automated neurological event detection has made possible this emerging engineering field, with new technologies and applications being created on an ongoing basis (Wolpaw

and Wolpaw, 2012).

However, as typical EEG signals are of the order of microvolts (μV), contamination by non-cerebral signals is frequent. These artefacts can significantly distort the EEG signal, making its interpretation difficult, and can dramatically disimprove automatic neurological event detection classification performance. In particular, contamination of EEG signals by artefacts arising from head movements have been a serious obstacle in the deployment of automatic neurological event detection systems in ambulatory EEG, i.e. environments where the patient or user has unrestricted movement. Similarly, analysis of epileptic and neonatal seizure detection systems developed by the Biomedical Signal Processing Group at University College Cork (UCC), have identified movement, ocular and respiratory artefacts as problematic, leading to a large number of false detections, and effectively preventing these automatic neurological event detection systems from being deployed in a clinical setting. This thesis, therefore, investigates and develops a number of promising artefact detection and removal algorithms for use in these automatic neurological event detection systems.

1.2 Aim and scope of this thesis

A significant portion of the artefact detection and removal techniques outlined in the literature address EEG recorded in environments where the user is instructed to limit behaviours that may cause artefacts. Alternatively, many papers validate the performance of their artefact processing algorithms on simulated or selective datasets. In order to effectively implement a real-world, automated neurological event-detection system, the algorithm must be capable of dealing with the entire range of EEG signals that may arise in noisy, medical environments. These systems must provide methods that comfortably deal with the widespread occurrence of EEG artefacts and be robust in the classification of the neurological event for which it is designed.

There are many papers published outlining methods of treating EEG artefacts (Barlow, 1984; Anderer et al., 1999; van de Velde et al., 1999; Delorme et al., 2001; James and Gibson, 2003; Gasser et al., 2005; Shoker et al., 2005; Winkler et al., 2011). An overwhelming majority of these papers focus on "cleaning" the EEG for visible inspection, often for one particular artefact type. However, fewer papers focus on those artefacts which are most problematic for specific event detection applications in a medical setting.

While many techniques are useful for visibly removing artefact by "cleaning" the EEG, comparatively few record how these artefacts affect the event detection systems in question. In many cases, this leads to artefact removal techniques which negatively impact on the performance of the event detections system by mistakenly removing valuable neural EEG information.

This thesis aims to explore a number of methods of improving two automatic neurological event detection systems under development at the Biomedical Signal Processing Group at U.C.C. The first of these is an epileptiform activity and epileptic seizure detection system for use in an ambulatory setting. This research was undertaken as part of the development of the REACT (**R**eal-time **E**EG **A**nalysis for Event Dete**CT**ion) ambulatory real-time EEG system and is based on data from routine EEG recorded at Cork University Hospital (CUH). The second event detection system investigated is an automated neonatal seizure detection system for real-time ward monitoring of at-risk babies in the neonatal intensive care nit (NICU). This research was carried out with the Neonatal Brain Research Group in developing the ANSeR (**A**lgorithm for **N**eonatal **Se**izure **R**ecognition) system and implemented with data collected at NICUs at Cork University Hospital and University College London Hospitals (UCLH).

1.3 Outline of thesis

The thesis is divided into the following chapters:

Chapter 2 introduces the different types of EEG artefacts; their causes, characteristics and morphologies. A comprehensive literature review of existing artefact detection and removal techniques is then presented. Additional literature reviews of EEG diagnostic systems in the form of automated epileptiform activity and epileptic seizure detection as well as automated neonatal seizure detection are also included. This chapter goes on to discuss existing methods of dealing with artefacts in these automated neurological event detection systems. Finally, this chapter introduces the datasets used to develop artefact detection and removal techniques in the thesis.

Chapter 3 aims to characterise the artefacts that arise from head-movements in an ambulatory environment. An artefact detection system is then designed to detect artefacts arising from head movements. A robust system is produced whereby artefacts arising from head-movements are accurately detected in the EEG.

Chapter 4 investigates the feasibility of using additional physical signals to detect EEG artefacts arising from head movements. Gyroscope signals from the Emotiv EPOC headset are used as a means of quantifying head-movement, and subsequently utilised to determine whether or not EEG artefacts have been produced in the EEG.

Chapter 5 explores the ways in which the EEG and gyroscope signals can be combined in order to improve the overall performance of head-movement artefact detection. A comprehensive multimodal, data fusion analysis of these signals at the feature and classifier levels is carried out.

Chapter 6 investigates artefact detection in the context of an automated neurological event detection system. Automated artefact detection is incorporated into an automated epileptiform activity detection system by performing classifier fusion of epileptiform and artefact detection classifiers. In doing so, false detections of epileptiform events are significantly reduced, resulting in state-of-the-art epileptiform activity detection when all (or almost all) epileptiform events must be detected.

Chapter 7 explores the issue of artefact removal from EEG. Many current technologies advocate the use of artefact removal techniques, either by detecting artefact and then rejecting the contaminated sections of EEG or by using some form of blind source separation technique to decompose the EEG into estimated source signals and remove those sources deemed artefactual. In order to investigate artefact removal in the context of automated neurological event detection systems, this chapter focuses on respiratory artefact in neonatal EEG. In the neonatal seizure detection system designed by the neonatal brain research group at UCC, respiratory artefact was found to be responsible for over 50% of false positive detections. Work in this chapter, therefore investigates the removal of respiratory artefact with blind source separation algorithms aided by additional physiological signals.

Chapter 8 summarises the main findings of this thesis, and places the findings in the context of current artefact detection and removal techniques. It goes on to outline the implications for EEG in general as well as automatic medical EEG diagnostic systems in particular. This chapter then outlines questions and issues resulting from this thesis and some future research directions are discussed.

1.4 Publications arising from thesis

Publications - Journals

O'Regan S., Faul, S. and Marnane W.P.,
"Automatic detection of EEG artefacts arising from head movements", *Medical Engineering & Physics*, 35(7),867-874, 2013

O'Regan S. and Marnane W.P.,
"Multimodal detection of head-movement artefacts in EEG", *Journal of Neuroscience Methods*, 218(1):110-120, 2013

O'Regan S., Kelleher D., Temko A., McNamara B., Costello D., and Marnane W.P.
"Robust automated epileptiform activity detection in the presence of ocular and movement artefacts", *In Preparation*

O'Regan S., Lightbody G., Temko A., Mathieson S., Boylan G. and Marnane W.P.,
"Respiration artefact removal in automated neonatal seizure detection using blind source separation and an electrocardiogram derived respiratory signal", *In Preparation*

Kelleher D., Temko A., **O'Regan S.**, Nash D., McNamara B., Costello D., and Marnane W.P.
"Epileptiform activity detection using support vector machines", *In Preparation*

Publications - Peer-reviewed Conferences

Kelleher D., Temko A., **O'Regan S.**, Nash D., McNamara B., Costello D., and Marnane W.P.
"Parallel artefact rejection for epileptiform activity detection in routine EEG", *In Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC)*, 2011, pp 7953:7956

O'Regan S., Faul, S. and Marnane W.P.,
"Automatic detection of EEG artefacts arising from head movements", *Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC)*, 2010, pp 6353:6356

O'Regan S., Faul, S. and Marnane W.P.,

"Automatic detection of EEG artefacts arising from head movements using gyroscopes", *Proceedings of the 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL)*, 2010, pp 1:5

Chapter 2

EEG Artefacts

2.1 Introduction

This chapter introduces the concept of EEG artefacts and gives an overview of the published literature that outlines their various characteristics and morphologies. A literature review of the signal processing and machine learning methods used in detecting and removing EEG artefacts is then presented. An overview of the published literature in automated epilepsy and neonatal seizure detection systems is subsequently outlined with particular emphasis on the role played by artefact processing in these environments. Finally, a detailed account of the data used in this thesis is presented.

2.2 The electroencephalogram

The electroencephalogram (EEG) can be broadly described as the recording of averaged electrical activity at different positions on the head. More specifically, EEG measures ionic current flows during synaptic excitation of the dendrites of large groups of pyramidal neurons in the cerebral cortex (Lopes da Silva, 1991). In scalp EEG, signals are measured between pairs of electrodes placed in a symmetrical array on the scalp (Figure 2.1). The most common standard placement guide for electrodes is the 10-20 system (Jasper, 1958), where each electrode site is identified by a letter and a number. The letters refer to the **F**rontal, **C**entral, **P**arietal, **T**emporal and **O**ccipital lobes of the brain. The numbers refer to the hemisphere locations; even numbers refer to the right hemisphere and odd numbers

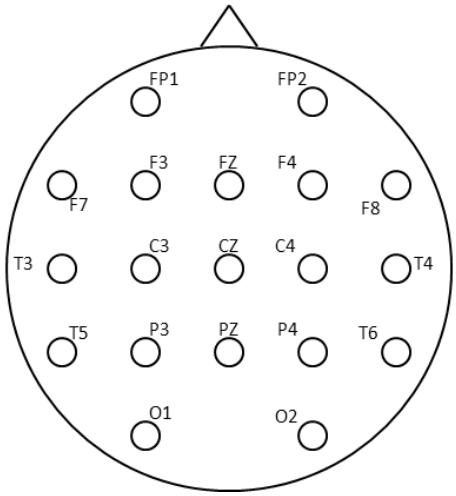


Figure 2.1: EEG electrodes placed according to the 10-20 system.

to the left hemisphere. 'Z' refers to an electrode placed on the mid-line. The number of EEG electrodes can vary with usage, usually ranging from 9 electrodes for neonatal monitoring to as many as hundreds of electrodes in epilepsy, evoked potential and brain computer interface applications. The amplitude of the EEG signal strongly depends on the level of synchronisation of the underlying neurons and is attenuated by thick layers of tissue (fluid, bone and skin). Typical EEG amplitudes are of the order of microvolts (μV). Examples of adult and neonatal EEG are displayed in Figure 2.2. Although adult and neonate EEG may appear similar to the untrained observer, they are considerably different in composition (Volpe, 2008), with the neonatal EEG exhibiting more complex behaviour than that of adult EEG. However, neurological event detection systems designed with adult EEG have been shown to work effectively with neonatal EEG (Gotman et al., 1997) and vice versa (Faul et al., 2009).

2.3 What are EEG artefacts?

An EEG artefact is any electrical potential appearing on the EEG trace that arises from a source other than the brain (Sanei and Chambers, 2007). Artefacts are prevalent in EEG recordings due to the low potential difference of EEG signals ($\sim \mu V$), which can easily be swamped by electrical noise from non-cerebral sources. However, the definition of what constitutes an EEG artefact can vary slightly with application. Most notably, in neurological event detection systems, some authors extend artefact status to signals emanating from cerebral sources other than those under investigation, e.g. Mitra et al.

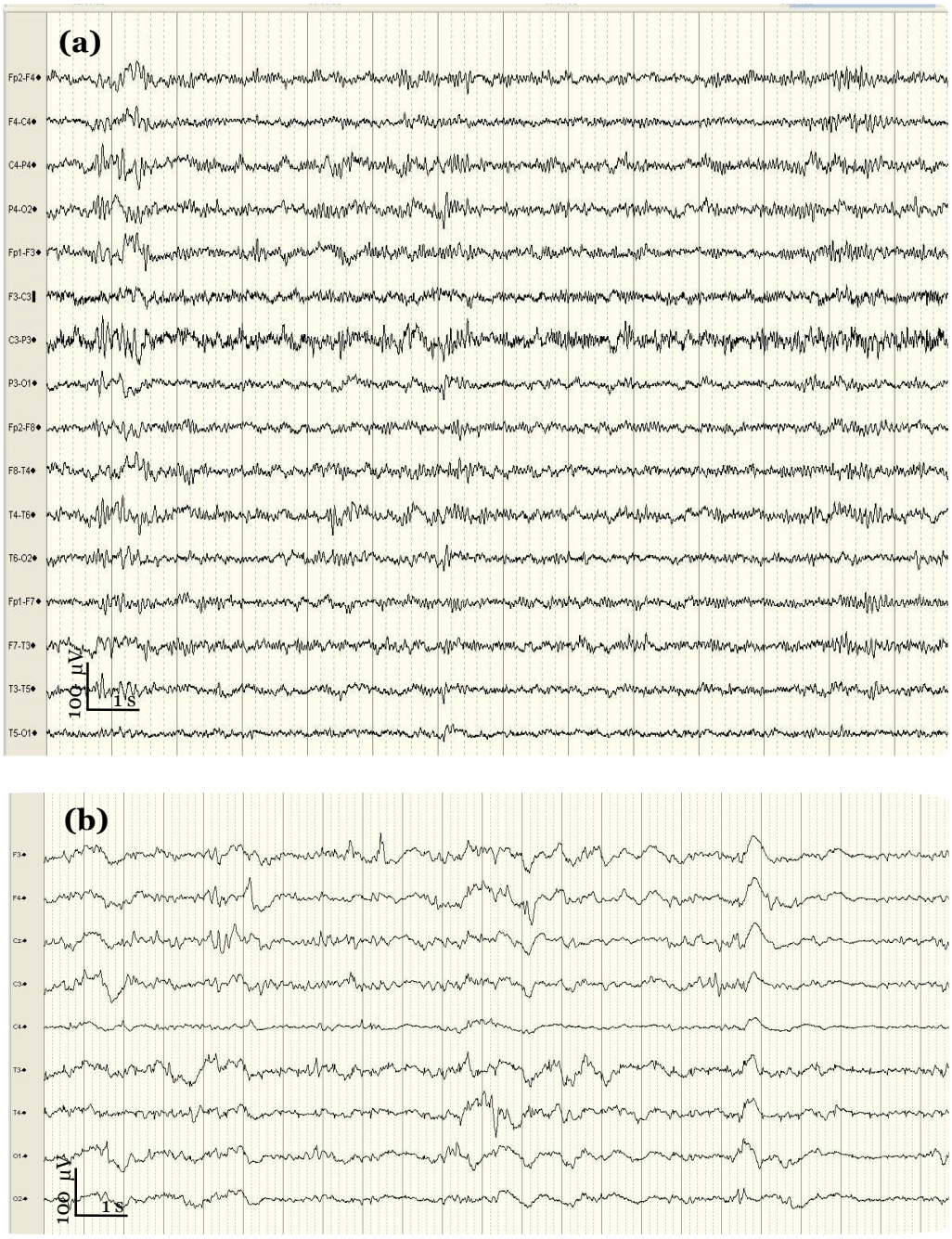


Figure 2.2: Examples of normal background EEG for an adult patient (a) and a neonatal patient (b). (a) Normal background adult EEG recorded using a Viasys Nicolet EEG machine in the Department of Neurology at Cork University Hospital. The displayed EEG is sampled at a frequency of 250 Hz. (b) Normal background neonatal EEG recorded using a NicOne EEG machine in the neonatal intensive care unit at Cork University Hospital. The EEG is sampled frequency of 256 Hz.

(2009) deem bi-frontal delta activity to be artefact in their neonatal seizure detection system. Conversely, other authors do not include certain non-cerebral, electrical signals in the artefact class, e.g. in sleep stage classification, electrocular potentials can be used to signify REM sleep and are thus deemed non-artefact. In this thesis, however, the most common definition of artefact will be employed; namely, an EEG segment will be deemed artefact if it arises from a source external to the brain. These artefacts can be broadly divided into two sub-categories: physiological and non-physiological. Physiological artefacts arise from non-cerebral, electrical signals within the body. Non-physiological artefacts originate from electrical sources external to the body.

2.3.1 Physiological artefacts

Ocular artefacts

Artefacts produced by vertical and lateral eye movements are extremely common in the EEG. The eyeball acts as an electrical dipole with a positive pole located at the cornea and a negative pole positioned at the retina (Picton et al., 2000). When the eyeball rotates about its axis, it generates a large-amplitude alternate current field, which is detectable by any electrodes near the eye. The electrodes that detect the ocular phenomenon most prominently are the ones that are closest to the eyeballs, namely the frontal electrodes: FP1, FP2, F7, and F8.

Vertical eye movements are typically observed with blinks. A blink causes the positive pole (at the cornea) to move closer to frontopolar (FP1-FP2) electrodes, producing symmetric downward deflections. During downward eye movement the positive pole (at the cornea) moves away from frontopolar electrodes, producing an upward deflection best recorded in channels FP1 and FP2.

Lateral eye movements mostly affect lateral frontal electrodes F7 and F8. During a left lateral eye movement, the positive pole moves toward F7 and away from F8. Electrodes FP1 and FP2 remain largely unchanged. Examples of eye blink and lateral eye movement artefacts are displayed in Figure 2.3.

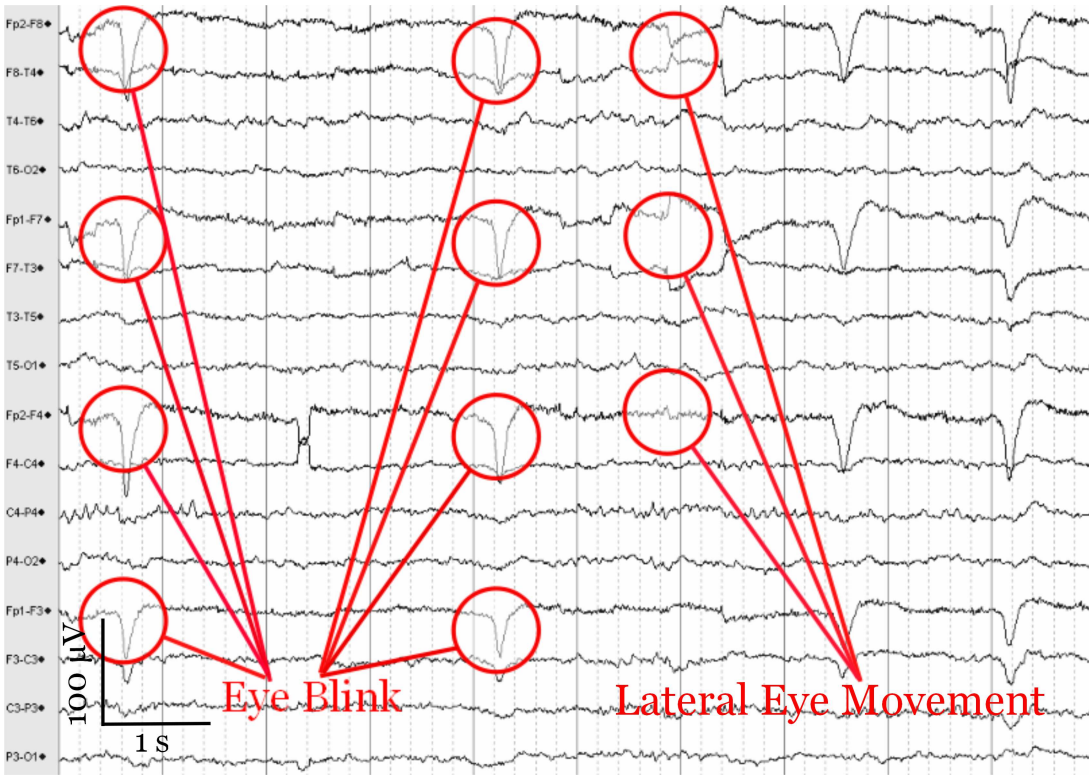


Figure 2.3: EEG sample showing a number of instances of eye blinks and lateral eye movements. The EEG was recorded using a Viasys Nicolet EEG machine in the Department of Neurology at Cork University Hospital. The displayed EEG is sampled at a frequency of 250 Hz.

Muscle artefacts

Myogenic potentials due to the firing of muscles located on or close to the scalp are a major cause of artefact in EEG. Frontalis and temporalis muscles, activated with clenching of the jaw or raising of the eyebrows, are common causes. Muscle or electromyogram (EMG) artefacts are predominantly high frequency signals whose amplitudes can vary depending on the muscle in questions as well as the strength of muscle firing (Goncharova et al., 2003). An example of EMG artefact arising from head movement is displayed in Figure 2.4.

Cardiac artefacts

Cardiac artefacts arise from the electromagnetic field produced by the heart and can take two forms: electrocardiogram (ECG) artefacts and pulse artefacts. ECG artefacts are caused by the electromagnetic field of the heart, whose potential varies over the surface

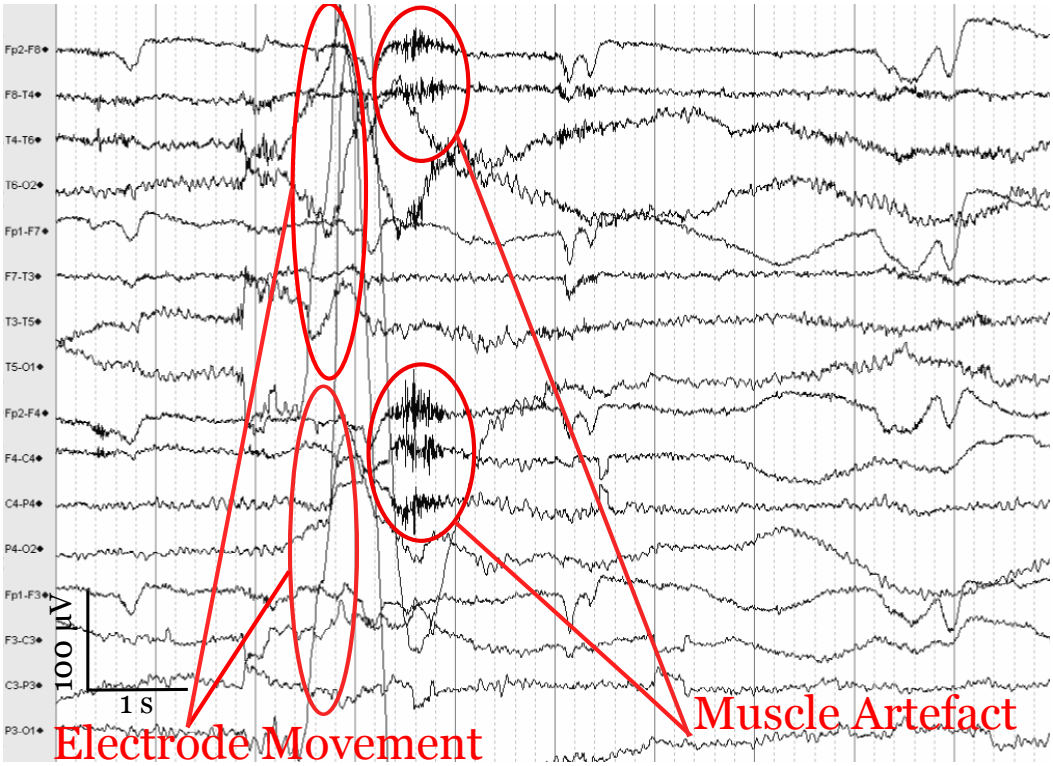


Figure 2.4: EEG sample showing artefacts arising from head movements.The EEG was recorded using a Viasys Nicolet EEG machine in the Department of Neurology at Cork University Hospital. The displayed EEG is sampled at a frequency of 250 Hz.

of the scalp. ECG artefacts often show up in the EEG as single, regularly spaced sharp waves, that correspond with the QRS complex of the ECG channel, and typically range in size from three hundred times smaller than the EEG to sizes comparable to the EEG signal (Devuyst et al., 2008). An example of ECG artefact appearing on a loose electrode is displayed in Figure 2.5. Pulse or ballistocardiogram artefacts occur when an EEG electrode is placed directly over a pulsating artery, becoming prominent if the electrode is loosely applied. The pulsation can cause slow waves that may simulate EEG activity, such as that displayed in Figure 2.6. Like ECG artefacts, there exists a direct relationship between the timing of ECG and the pulse waves appearing on the EEG; the QRS complex slightly precedes the pulse waves.

Respiration artefacts

Respiration artefacts occur due to movement of an electrode with inhalation or exhalation. This can manifest itself in one of two ways: the first of these is a respiration artefact taking the form of slow, rhythmic EEG activity, synchronous with the body movements associated

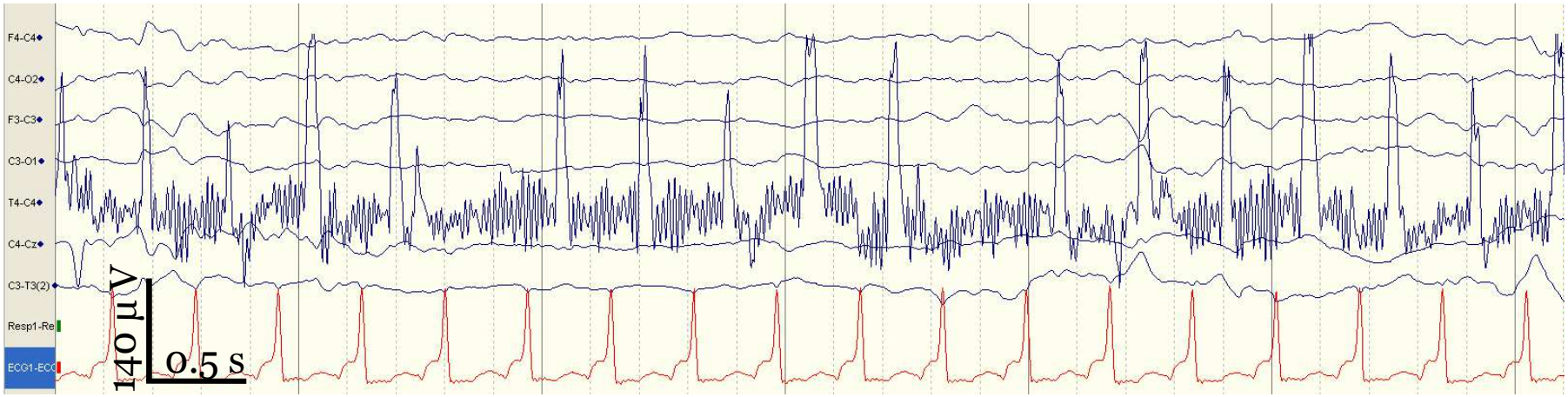


Figure 2.5: Loose electrode on channel T4-C4, resulting in high frequency noise and pulse artefact. The ECG trace (red) shows the QRS complex preceding the ECG by 200-300 ms.

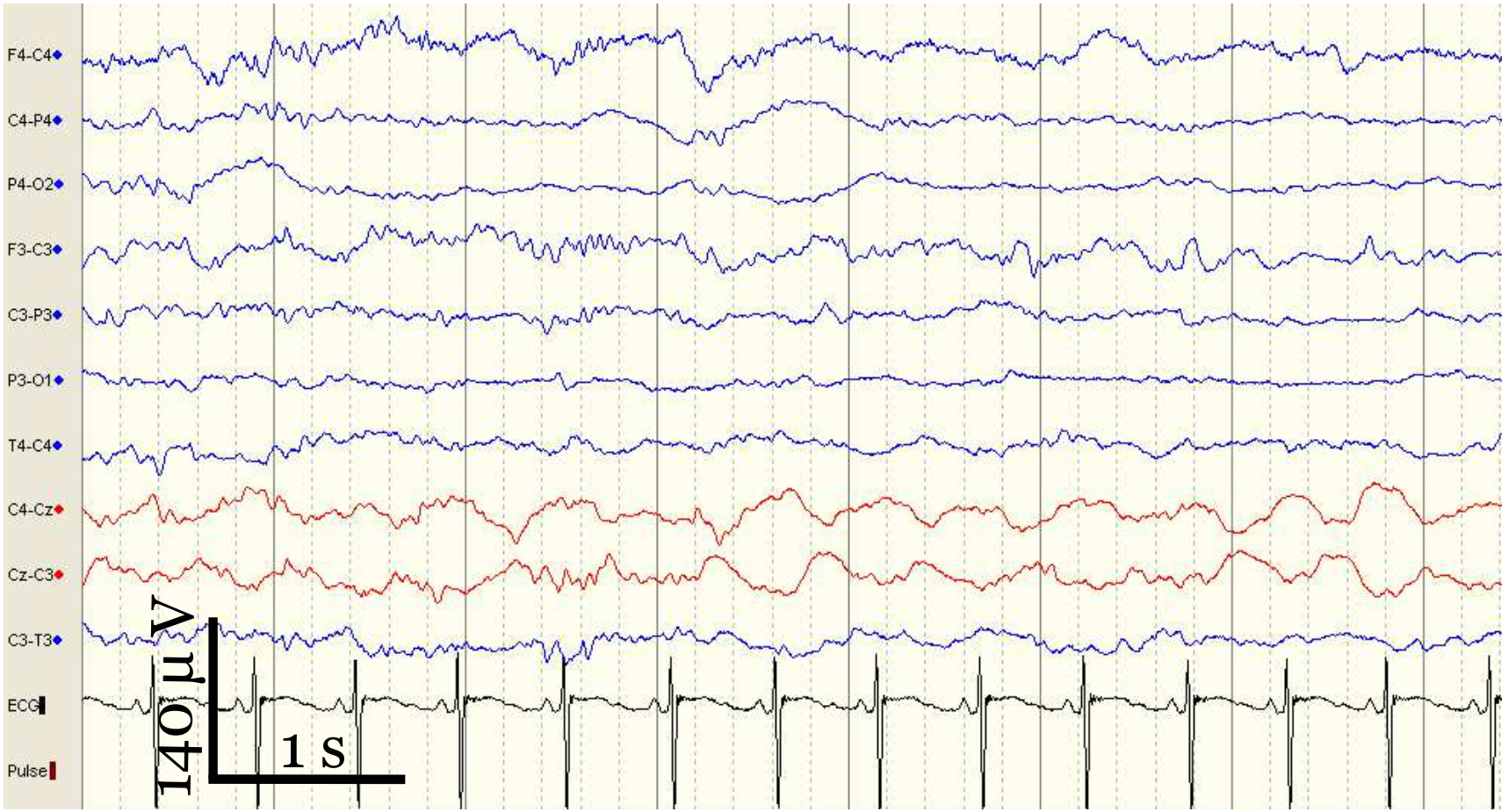


Figure 2.6: Pulse artefact on neonatal EEG recording; visible on channels C4-Cz and Cz-C3 (red) manifesting as a slow wave. The ECG trace (black) shows the QRS complex preceding the pulse artefact by 200-300 ms.

with breathing and mechanically affecting the impedance of (usually) one electrode. An example of slow-wave respiration artefact is displayed in Figure 2.7, where electrode C4 has recorded respiration artefact that consequently appears on channels F4-C4 and T4-C4, and to a lesser extent on channels C4-P4 and C4-Cz. A second form of respiration artefact, which is especially common in neonatal EEG, can be slow or sharp waves (such as that displayed in Figure 2.8) that occur synchronously with inhalation or exhalation and involve those electrodes on which the patient is lying. This form of respiration artefact can often mimic neonatal seizure.

Electrodermal artefacts

Electrodermal or sweat artefacts can originate from changes in the electrolyte concentration of the EEG electrodes due to sweat secreted from eccrine sweat glands on the scalp. The artefacts occur when sodium chloride and lactic acid from sweat react with the metals of the electrodes. Sweat artefacts take the form of long, slow baseline sways, often in the frequency range of 0.25 - 0.5 Hz. An example of electrodermal artefact is displayed in Figure 2.9, where slow baseline drifts can be observed on channels F4-C4, F3-C3, and C3-T3.

Glossokinetic artefacts

The tongue, like the eyeball, acts as an electric dipole, with the tip negative with respect to the base. As the tip of the tongue moves relative to the base, this reverses the electric field, and can cause a deflection in the EEG (Vanhatalo et al., 2003). Similar to the production of ocular artefacts, the glossokinetic dipole produces a broad potential field that drops from frontal to occipital regions. The amplitude of glossokinetic artefacts tends to be significantly smaller than ocular artefacts however, with the drop in potential away from frontal electrodes being less steep for glossokinetic than ocular artefacts. These artefacts can be introduced by talking, in particular with the use of the letter ‘l’. Chewing and sucking can also produce glossokinetic artefacts, often accompanied by EMG artefacts.

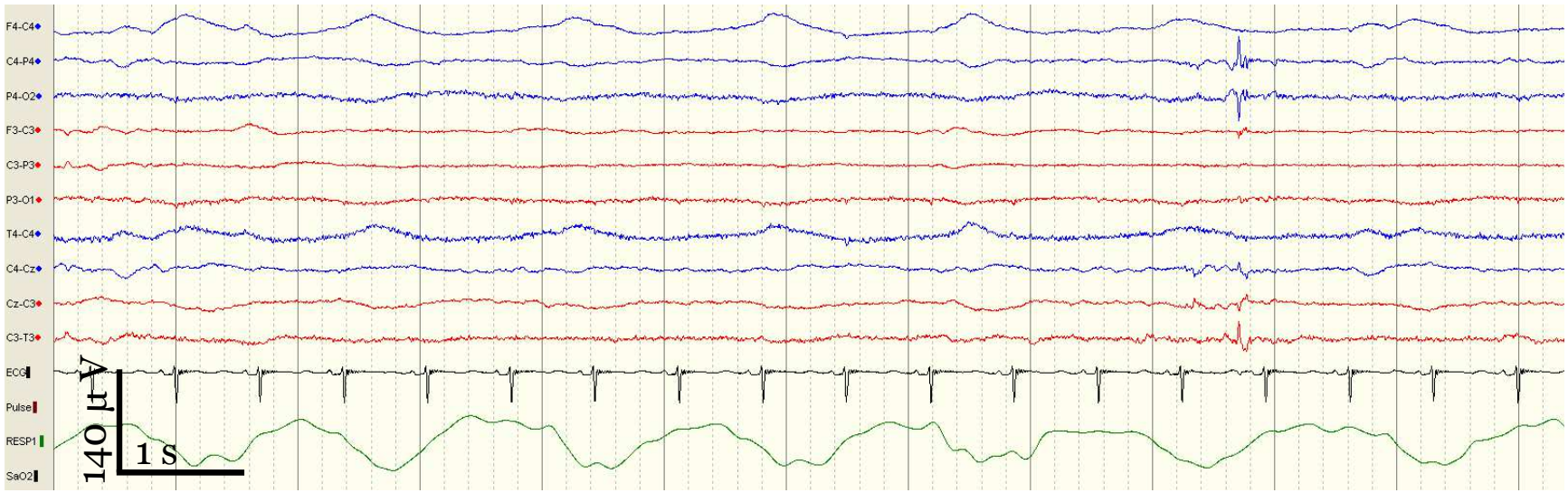


Figure 2.7: Respiration artefact, manifesting as slow waves, on neonatal EEG recording; electrode C4 records the artefact which is then visible on channels F4-C4 and T4-C4, and to a lesser extent on channels C4-P4 and C4-Cz.

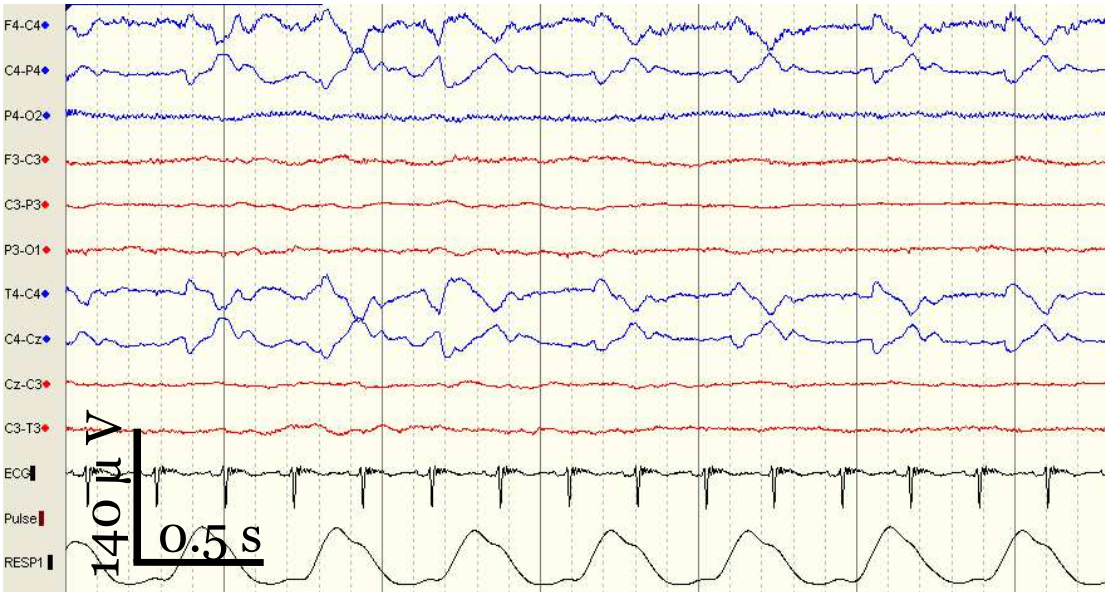


Figure 2.8: Respiration artefact on neonatal EEG recording; visible on channels F4-C4, C4-O2, T4-C4 and C4-Cz, manifesting as slow and sharp waves.

2.3.2 Non-physiological artefacts

Mains voltage

A 50 or 60 Hz artefact may appear on the EEG line due the mains voltage. This artefact can be transferred directly via power lines if the EEG is not adequately grounded. Alternatively, if the impedance of one of the active electrodes becomes large with respect to the ground electrode, the ground electrode can act as an active electrode, introducing mains voltage artefact into the EEG (Tatum et al., 2011). Mains voltage artefact may also appear on the EEG via the use of fluorescent lights or other electrical equipment in the close vicinity of the EEG machine.

Electrode pop and electrode movement

Another common electrode artefact is electrode pop, which occurs with momentary loss of contact between the electrode and the scalp. This causes an abrupt impedance change, morphologically appearing as single or multiple sharp waveforms. These sharp waveforms are typically high amplitude vertical transients which are usually confined to a single electrode. Electrode movement occurs when the electrode moves with respect to the scalp. These movements can produce high-amplitude deflections in the EEG of the order

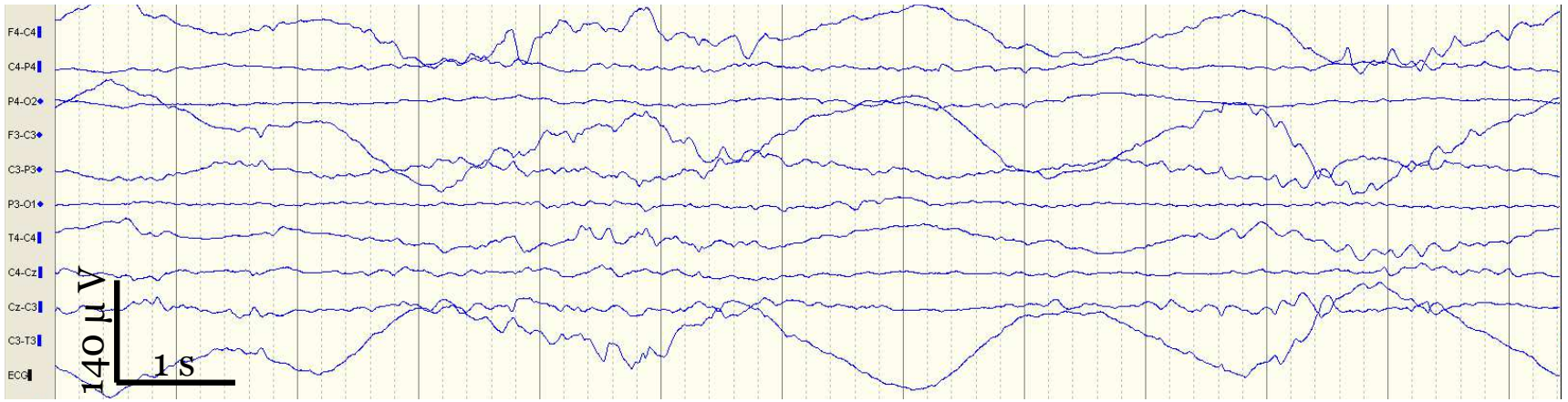


Figure 2.9: Electrodermal artefact on neonatal EEG recording, most recognisable by the slow, baseline drifts on channels F4-C4, F3-C3, and C3-T3.

of milliVolts. Unlike, electrode pop, electrode movements deflections are slower, usually in the range of 1 - 10 Hz.

Loose electrode

An electrode that is not making satisfactory contact with the scalp, can lead to an impedance change on the electrode manifesting as a slower, more prolonged EEG spike than that seen during electrode pop.

Machine artefacts

Various medical devices in the ICU can cause artefacts in the EEG. In particular, the use of automatic electric infusion pumps introduces infusion motor artefact (IMA). Morphologically, IMA appears as very brief spiky transients, sometimes followed by a slow component of the same polarity. Artefacts arising from the use of a gravity-fed intravenous infusion may also manifest itself in the EEG signal (Egol and Guntupalli, 1983). This artefact, which is thought to be due to electrostatic charges on the infusion drops, appears as spike-transient potentials at fixed intervals that coincide with drops of the infusion. Ventilator equipment may also introduce artefacts into the EEG (Tatum et al., 2011). This artefact type is often related to the respiration signal, but may vary in morphology and frequency, based on the machine in question, the parameters of the machine and the individual. Finally, artefacts due to movement of other people in the vicinity of the EEG recording have been known to introduce electrostatic artefacts into the EEG, often resembling interictal epileptiform discharges (Tatum et al., 2011).

If EEG is recorded inside an fMRI machine, a gradient artefact arises due to the magnetic fields of the fMRI machine. These artefacts can be of the order of 50 times the amplitude of background EEG and are the focus of a growing research field (Grouiller et al., 2007).

2.4 Artefact avoidance and minimisation

In many EEG recordings, participants or patients are instructed to remain still and avoid where possible, excessive eye movements, blinks and head-movements. From a data loss and computational perspective, artefact avoidance is the most ideal method of treating

EEG artefacts, as the presence of artefacts in the data is minimised (Fatourechi et al., 2007). However, there are several drawbacks to employing this approach. Since many physiological signals (such as heart beats) are involuntary, some artefacts will always be present in EEG signals. Furthermore, even when instructed to limit head and eye movements, patients often have difficulty in doing so as it is not always easy to control these impulses, especially in longer recordings. This issue becomes even more pronounced in the case of EEG monitoring of neonates or children. As many event detection systems are recorded in the ICU, avoiding contamination by extra-physiological artefacts due to medical equipment and health care professionals is simply impossible. In an ambulatory setting, where it is intended that patients go about their daily lives as normal, artefact avoidance is similarly infeasible. This thesis, therefore, aims to employ artefact detection and removal techniques for those artefacts that remain in the EEG, particularly when artefact avoidance is impossible or unsuccessful. These techniques are drawn primarily from biomedical signal processing and machine learning.

2.5 Biomedical signal processing and machine learning

In this thesis, signal processing techniques are utilised in order to reduce the effect of artefacts on the EEG, both for the clinician and also for automated neurological event detection algorithms. Broadly speaking, artefact processing can be divided into two approaches: artefact detection and artefact removal. Artefact detection is concerned with mimicking a human EEG observer in indicating which sections of the EEG signal are contaminated with artefact; this process is automated with machine learning algorithms. Artefact removal aims to take advantage of differences in the properties of artefact and cerebral EEG to "clean" the EEG.

2.5.1 Machine learning

Learning from data is used in situations where there is no available analytical solution, but where there is data that can be used to construct an empirical solution. Machine learning then concerns itself with the automation of this process, i.e. "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" (Mitchell, 1997). In doing so, the performance should generalise such that the system

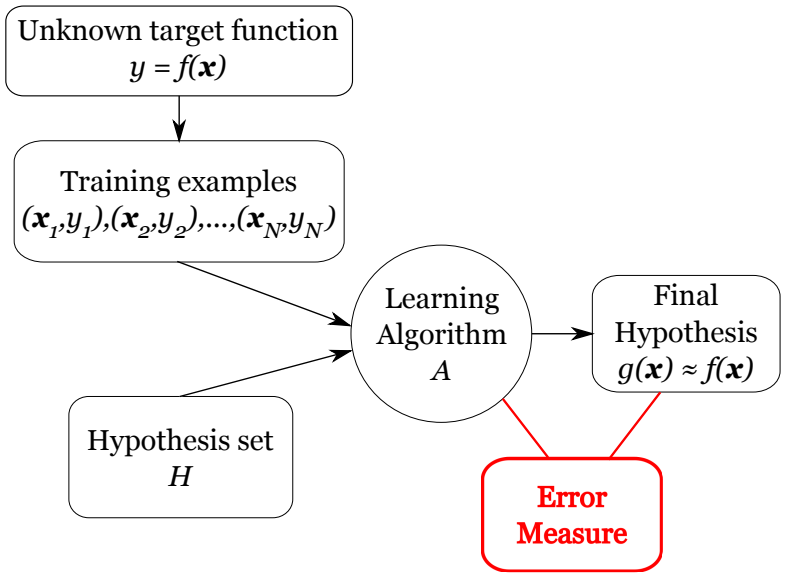


Figure 2.10: A basic representation of learning from data.

will perform well on unseen data instances. The learning process is illustrated in Figure 2.10, where training examples $(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)$ are used to select a suitable hypothesis $g(x)$ from a hypothesis set H via a learning algorithm A. The selection of the hypothesis $g(x)$ is chosen based on minimising an appropriate error measure E. In this thesis, the branch of machine learning known as *supervised learning* is used. In supervised learning, the training data set comes with explicit labels indicating the correct output for those data points. That is to say, the data set has been annotated so that a machine learning classifier knows when a training data point (features extracted from an epoch) belongs to a predefined class. The learning algorithm then utilises this information to select a (hopefully) optimal decision boundary such that additional, unlabelled test data points will be classified successfully.

Once artefactual sections of EEG have been identified with a machine learning algorithm there are a number of options available. Artefact EEG could be highlighted to the clinician as an automated artefact annotator, to aid in their decision making. Alternatively, sections of artefact EEG could be automatically rejected before the clinician examines the data, or before the data is sent to an automated neurological event detection system. This approach is particularly common in evoked or event-related potentials applications and is known as conventional trial rejection (CTR) (Luck, 2005). Finally, the artefact detector could be utilised in conjunction with another classifier designed to detect some neurological event of interest such as epileptic seizure or imagined left/right arm movement as used in BCI applications.

In artefact detection and rejection, sections of EEG containing artefact are either highlighted (detection) or entirely discarded (rejection). Owing to the nature of the EEG recording process, in practice these artefact-ridden sections of EEG will comprise both cerebral EEG and artefact signals resulting in a loss of data. As EEG electrodes measure the electrical activity on the scalp, each electrode measures the superposition of activity from several underlying sources. As discussed, often these electrical sources originate external to the brain (artefacts), and therefore are of no interest in the context of EEG measurements. Artefact removal thus aims to reverse this process, by removing only that portion of the signal attributed to artefact, while leaving the cerebral EEG largely intact. Artefact removal techniques can be broadly split into two categories: (i) filtering and regression and (ii) blind source separation (more recently referred to under the umbrella-term linear instantaneous mixture models).

Although technically speaking, all artefact removal methods are filtering of the EEG (in the time, frequency or spatial domains), filtering here will refer to those time-frequency methods based upon the Fourier transform. In these methods, it is assumed that the frequency of artefact differs substantially from the cerebral EEG. The relevant artefactual frequency ranges are then assumed to be fixed (linear filtering), or are determined from additional physiological signals (adaptive filtering and regression).

2.5.2 Blind source separation

Blind source separation (BSS) is a family of statistical models corresponding to coordinate transformation in data space. On multivariate data, such as EEG, information from several variables (i.e. channels) can be used to construct more insightful new variables by applying a coordinate transformation to the data (Comon and Jutten, 2010). In the context of artefact removal, this can translate to the separation of the contribution of cortical signals and artefact. The EEG measurement signals $\mathbf{X}$ can thus be modelled as a linear transformation of latent variables $\mathbf{Y}$ corresponding to the actual underlying cortical signals as well as electrical signals arising from non-cerebral sources, such as heart, eyes or muscle tissue. This linear transformation of the N-channel EEG is therefore modelled by an unknown, full-rank, *mixing matrix*, $\mathbf{A}$, such that:

$$\mathbf{X} = \mathbf{AY}, \tag{2.1}$$

where $\mathbf{A} \in \mathbb{R}^{N \times N}$. In the presence of samples arising from $\mathbf{X}$, the goal of BSS is therefore to estimate an inverse of $\mathbf{A}$ that allows the accurate recovery of the latent variables $\mathbf{Y}$. In other words, find the transformation $\mathbf{A}^{-1}$ to coordinates corresponding to the underlying cortical signals as well as artefact sources.

To find a suitable *de-mixing* matrix $\mathbf{A}^{-1}$ requires further assumptions on the statistical properties of the latent variables $\mathbf{Y}$ as well as assumptions on $\mathbf{A}^{-1}$. Differing assumptions allows different coordinate transformations and gives rise to a plethora of available BSS algorithms, among which independent component analysis (ICA), principal component analysis (PCA) and canonical correlation analysis (CCA) are perhaps the most commonly used. Due to the prominence of ICA-based methods in the literature, the following paragraphs provide a brief introduction to independent component analysis; for simplicity, in each case it is assumed that the number of sources M is equal to the number of channels N. While in many EEG recording instances M may not in fact be equal to N James and Hesse (2005) shows that this assumptions performs sufficiently for most scenarios and where $M > N$, a pre-processing dimensionality reduction step can be performed (typically with PCA).

Independent component analysis

Independent component analysis (ICA) is a computational method for separating a multivariate signal into additive subcomponents that supposes the mutual statistical independence of the non-Gaussian source signals. ICA can be split into two main steps: (i) whitening of the EEG observations so that signals are normalised and uncorrelated and (ii) an orthogonal transformation of the whitened signal so as to maximise independence between the signals.

The first step in ICA is that of *whitening*, i.e. multiplication of the EEG signals $\mathbf{X}$ by a whitening matrix $\mathbf{W}$ such that:

$$\mathbf{Z} = \mathbf{X}\mathbf{W}, \tag{2.2}$$

and,

$$\mathbf{W} = cov(\mathbf{X})^{-1/2}. \tag{2.3}$$

Whitening ensures that the covariance matrix of $\mathbf{Z}$ is the identity, and therefore that the signals are uncorrelated and normalized.

The second step in ICA then is an orthogonal transformation (i.e. a transformation that leaves the covariance unchanged) of $\mathbf{Z}$ so as to maximise the independence between signals:

$$\mathbf{Y} = \mathbf{R}\mathbf{Z}, \tag{2.4}$$

where $\mathbf{R}\mathbf{R}^T = \mathbf{I}$. Maximising the independence of sources corresponds to minimising the mutual information,

$$MI = \sum_{j=1}^{M} H(\hat{\mathbf{y}}_j) - H(\hat{\mathbf{Y}}), \tag{2.5}$$

between estimated sources, where $H(.)$ is the entropy, $\hat{\mathbf{Y}}$ are the estimated sources and $\hat{\mathbf{y}}_\mathbf{j}$ are the columns of $\hat{\mathbf{Y}}$. Finding the optimal rotation matrix $\mathbf{R}^*$ to give the maximally independent source estimates $\hat{\mathbf{Y}}$ becomes an optimisation problem such that:

$$\mathbf{R}^* = \underset{\mathbf{R}\mathbf{R}^T=I}{\operatorname{argmin}} \quad MI\left(\hat{\mathbf{y}}_1, ..., \hat{\mathbf{y}}_M\right) \tag{2.6}$$

$$= \underset{\mathbf{R}\mathbf{R}^T=I}{\operatorname{argmin}} \quad \sum_{j=1}^{M} H(\hat{\mathbf{y}}_j) - H(\hat{\mathbf{Y}}), \tag{2.7}$$

Since $\hat{\mathbf{Y}}$ is obtained from the observed EEG signals $\mathbf{X}$ by the linear transformation $\mathbf{R}$:

$$\mathbf{R}^* = \underset{\mathbf{R}\mathbf{R}^T=I}{\operatorname{argmin}} \quad \sum_{j=1}^{M} H(\hat{\mathbf{y}}_j) - H(\mathbf{X}) - \log\det\mathbf{R}, \tag{2.8}$$

and since the entropy of $\mathbf{X}$ does not depend on $\mathbf{R}$ and $\det\mathbf{R} = 1$ (as it is a rotation):

$$\mathbf{R}^* = \underset{\mathbf{R}\mathbf{R}^T=I}{\operatorname{argmin}} \quad \sum_{j=1}^{M} H(\hat{\mathbf{y}}_j). \tag{2.9}$$

Of all distributions with a given variance, the Gaussian distribution is the one with the highest entropy. Negentropy J measures the difference in entropy between a given distribution and the Gaussian distribution with the same variance. Thus, it follows from the definition of negentropy that:

$$\mathbf{R}^* = \underset{\mathbf{R}\mathbf{R}^T=I}{\operatorname{argmin}} \quad \sum_{j=1}^{M} \Big(H\big(\mathcal{N}(\mu(\hat{\mathbf{y}}_j), cov(\hat{\mathbf{y}}_j))\big) - J(\hat{\mathbf{y}}_j) \Big). \tag{2.10}$$

As the entropy of a Gaussian distribution is invariant under the rotation $\mathbf{R}$,

$$\mathbf{R}^* = \underset{\mathbf{R}\mathbf{R}^T=I}{\operatorname{argmin}} \quad \sum_{j=1}^{M} -J(\hat{\mathbf{y}}_j). \tag{2.11}$$

In summary, minimising the mutual information is equivalent to maximizing the distance to a Gaussian distribution of the individual sources, as measured by the negentropy. This optimisation can then be achieved by gradient descent (Bell and Sejnowski, 1995). In addition to the approach outlined above, there are many alternative methods of implementing ICA based upon either direct minimisation of the mutual information or by maximisation of the non-Gaussianity of the sources: e.g. by approximate diagonalisation of the fourth-order cumulant tensor (kurtosis) (Cardoso and Souloumiac, 1993), or by deflation-type fixed-point algorithms (Hyvärinen and Oja, 1997).

2.6 Manual and semi-automatic artefact removal

The blind signal processing techniques mentioned in Section 2.5 were initially used to allow the clinician or researcher to manually remove artefacts from contaminated sections of EEG. These methods offer the possibility of removing artefacts while retaining the underlying cerebral portion of sections of EEG contaminated by artefact, thus reducing the potential for discarding clinically relevant EEG as takes place in CTR or linear filtering.

In epilepsy research in particular, a large body of literature has been published outlining semi-automatic methods of cleaning the EEG using blind source separation BSS techniques. In these methods, the multi-variate EEG signal is transformed into estimated source signals and a trained electroencephalographer can then identify and remove those components corresponding to non-cerebral sources. The remaining non-artefact sources can then be recombined to give a "cleaned" EEG signal. Following the introduction of applying independent component analysis to EEG data (Makeig et al., 1996; Vigário, 1997) an array of methods using ICA to remove artefacts from the EEG have been published. Delorme et al. (2001) introduced an artefact removal technique using ICA and higher order components. Nam et al. (2002) have used ICA to remove artefactual sources from ictal scalp EEG. While the bulk of blind source separation techniques applied to the EEG artefact removal problem have focused on ICA, a number of other blind source separation techniques have been investigated. Jung et al. (2000) compared the effectiveness of PCA and ICA in removing ocular artefacts from EEG and found that ICA was superior.

While utilising blind signal separation techniques represented a step forward in dealing with EEG artefacts, (in the sense that it is no longer necessary to completely discard any section of EEG containing artefact or lose significant EEG information via attenuation by filtering), the requirement of manually identifying and removing artefact sources is time-consuming and difficult. Consequently, attempts have been made to simplify this task for the researcher or clinician by providing information as to which sources likely correspond to artefact, or even to performing automated artefact removal in clear-cut cases. These methods are referred to here as semi-automated artefact removal techniques as they vastly reduce the complexity of the artefact removal task, albeit with some necessary input by the clinician/researcher. Iriarte et al. (2003) and Urrestarazu et al. (2004) implemented semi-automatic methods of removing EMG and ocular artefacts from ictal, scalp EEG with ICA. Qualitative results showed improvement in epileptic seizure clarity, as graded by a clinician. Liu et al. (2004) introduced a blind source separation technique for semi-automatic artefact removal for intra-cranial EEG on epileptic patients. A similar method was proposed by Campos Viola et al. (2009) to reduce ocular and ECG artefacts from EEG for event-related potentials applications.

Semi-automatic artefact removal techniques are not limited to ICA. Ille et al. (2002) outlined a semi-automatic method of using principal component analysis and spatial filters to remove ocular artefact in patients undergoing epileptic seizure. De Clercq et al. (2006) used canonical correlation analysis to create a semi-automatic method of removing muscle artefact from EEG; in that experiment, CCA outperformed linear filtering and a manual ICA method, in cleaning sections of EEG in patients with refractory partial epilepsy. De Clercq et al. (2006) proposed that the low autocorrelation of muscle artefact sources compared to ictal sources makes CCA attractive for its removal from ictal EEG. This method was repeated by Vergult et al. (2007) on a larger data set with similar qualitative results. Similarly, Crespo-Garcia et al. (2008) showed the usefulness of semi-automatic ICA and second order statistical (SOS) algorithms in removing muscle artefact for sleep research.

As these methods require a clinician/ researcher to help differentiate between cerebral and artefact sources of the decomposed EEG, they are unsuitable for use in automated neurological event detection algorithms. They do however highlight the fact that if artefact sources can be accurately identified, they offer much potential for automated neurological event detection systems. A review of methods that offer full automation are outlined in Section 2.7.2.

2.7 Automated EEG artefact detection and removal: A review

Artefact detection and removal techniques date back to the introduction of computerised EEG analysis in the 1970's. A wide range of techniques have been suggested in the literature, primarily in the areas of epilepsy, evoked and event-related potentials, brain-computer interface and sleep research. The earliest methods involved simple amplitude thresholding and linear filtering. More recently, advanced signal processing and machine learning algorithms have been applied with reasonable success. As mentioned in Section 2.5, these techniques can be broadly subdivided as follows: (i) automated artefact detection and rejection and (ii) automated artefact removal. The following section provides an overview of published attempts at automatically detecting and removing artefacts in the EEG.

2.7.1 Artefact detection

Thresholding methods

Many early artefact detection techniques advocated the rejection of any EEG section where the EEG amplitude exceeded a pre-defined threshold. These thresholding methods were often applied directly to the EEG signal; for example Barlow (1986) used thresholding of the EEG amplitude to remove data contaminated with electrode pop artefact. Rowland (1968) incorporated a similar thresholding method by using additional physiological signals (EOG and EMG), and rejected any EEG sections if the corresponding EOG or EMG signal exceeded a pre-determined threshold. Gevins et al. (1977) used thresholding of the EEG in different frequency bands to reduce ocular artefacts. A similar method was employed by Pfurtscheller et al. (1996) and McFarland et al. (1997) to reduce EMG artefact contamination in the EEG for BCI applications. An obvious disadvantage to these thresholding methods is that they do not allow the rejection of contaminated trials when EOG or EMG amplitude is small (Rowland, 1968; Croft and Barry, 2000). In evoked and event-related potentials research, Junghöfer et al. (2000) offered an alternative to simply neglecting channels that were contaminated by artefact by replacing rejected sections of one EEG channel with an aggregated EEG representation taken from nearby channels. Similarly, Mourad et al. (2007) proposed an alternative to rejecting high amplitude EEG signals, whereby a smoothing matrix was calculated to replace any high amplitude EEG.

However, these methods were introduced for high-density electrode placements and are insufficient if artefacts appear simultaneously on multiple channels and are unsuitable for recordings where electrode numbers are limited (such as in many clinical set-ups and in all neonatal EEG recordings).

Machine learning for artefact detection

Epoch-based artefact rejection systems segment the EEG into short windows and reject these epochs if artefact is detected within them; this CTR approach is especially popular in event-related potentials research. Fujioka et al. (2011) compared this method to independent channel rejection (He et al., 2007) and artefact blocking (Mourad et al., 2007), and found that rejecting epochs is suboptimal if the signal-to-noise ratio is low. In sleep research, Brunner et al. (1996) proposed a simple threshold-based classifier to detect EMG artefacts by comparing spectral EEG features to the same features extracted from background EEG. Similarly, Durka et al. (2003) used a simple classifier-based method of removing ocular and mains voltage artefacts in polysomnographic recordings. In this system, a number of features were identified as correlating with artefact, and a threshold was used to separate between classes. While statistical thresholding may be a reasonable first approach to artefact detection, especially for rejecting sections of highly contaminated EEG, single features do not provide sufficient information for distinguishing between artefact and non-artefact in most cases. This fact is taken into account by Durka and Blinowska (1996), who proposed a neural network classifier to identify EEG containing ocular artefacts. Results for classification of raw EEG performed below chance for test data; however, by pre-processing the EEG using wavelet analysis, 71 % of artefact epochs were correctly classified (true positives) and 82 % of normal EEG was correctly classified (true negatives). This method is promising in terms of its ability to accurately detect EEG artefacts; however, it has been demonstrated to detect ocular artefacts alone, in the absence of any other neurological event to be detected.

2.7.2 Artefact removal

Automated artefact removal techniques can be broadly split into two categories: (i) filtering and regression and (ii) blind source separation.

Filtering and regression

Accepted methods of removing non-physiological artefacts tend to be straightforward, with mains and machine artefact often removed by simple, linear filtering at the required frequency cut-off, and electrode pop and electrode movement, removed by rejecting sections of EEG where an amplitude threshold is exceeded. Due to the variability in the morphology between and within physiological artefact types, the majority of artefact removal techniques are tailored to removing physiological artefacts.

Linear filtering is useful for removing artefacts located within specific frequency bands that do not overlap with those of the neurological phenomena of interest and that do not vary with time (Barlow, 1984; Ives and Schomer, 1988). Accepted methods of removing non-physiological artefacts thus tend to be straightforward, with mains and machine artefact often removed by simple, linear filtering at the required frequency cut-off. Linear filtering was commonly used in early clinical studies to remove artefacts in EEG signals (Gotman et al., 1973). Traditionally, low-pass filters have been used to remove EMG artefacts and high-pass filters have been used to remove EOG and sweat artefacts as well as EEG baseline drift. The primary advantage of linear filtering is its simplicity. A secondary benefit is that additional physiological information in the form of EOG or EMG signals are not necessary to remove the artefacts. However, linear filtering fails when the neurological phenomenon of interest and artefact lie in similar frequency bands (de Beer et al., 1995). A look at the frequency ranges of neurological phenomena used in neonatal (0-32 Hz) and epileptic (0-64 Hz) EEG shows that for ocular (0-16 Hz) , muscle (10-100 Hz), cardiac (1-3 Hz) and respiration (0-12 Hz) artefacts this is usually the case (Volpe, 2008; Rowan and Tolunsky, 2003; Cacioppo et al., 2007). As a result, a simple filtering approach to remove EMG or EOG artefacts may lead to the loss of valuable neurological information in the process. A further drawback of using linear filtering to remove EMG artefact has been detailed by several authors who have described how low-pass filtering of EEG containing muscle artefacts may cause the filtered signal to closely resemble epileptic activity (Klass, 1995). Furthermore, linear filtering alone offers no solution to situations where the frequency range of the artefact changes over time, or where filtering is required only when artefact is present.

Multimodal regression and *adaptive filtering* using a linear combination of the artefact-contaminated EEG signal and an artefact reference signal are common methods for use in removing ocular and cardiac artefacts. If a reference signal for an artefact is available

(and the morphology of that reference signal is similar to that of the EEG artefact), it can be removed from the EEG, either by subtraction after scaling it by an appropriate factor (regression in the time domain) or by adapting the filter parameters based on the additional physiological signal (adaptive filtering and regression in the frequency domain). In this manner, ocular artefacts were removed from the EEG using regression with the EOG signal in either the time domain (Gratton et al., 1983) or the frequency domain (Whitton et al., 1978). A comparison between several EOG minimization methods based on regression analysis in time and/or frequency domain can be found in Jervis et al. (1988) and Brunia et al. (1989). Croft and Barry (2000) offer a more up-to-date comparison of ocular artefact removal techniques. Since then, Schlögl et al. (2007) produced a widely-cited, fully automated ocular removal system using regression with the EOG signal. However, in the absence of dedicated EOG channels, this approach is clearly not feasible.

Linear regression becomes more challenging for EMG artefacts, since they have no single suitable reference channel (Barlow, 1986) and applying regression using signals from multiple muscle groups requires multiple reference channels (Jung et al., 1998). Removal of cardiac artefacts poses a different problem, in that the artefacts often appear morphologically different to the recorded ECG signal. To deal with this, Strobach et al. (1994) suggested using the ECG reference signal as a trigger to an artificial ECG artefact reference model signal to remove ECG artefact from the EEG using linear regression. When an ECG spike was detected, an ECG artefact signal was generated from a reference model, scaled to the detected ECG signal and was then subtracted from the EEG signal. Sahul et al. (1995) presented an adaptive noise canceller for ECG artefact suppression in sleep EEG, reporting slowly changing filter weights over the night.

An inherent drawback of all artefact removal techniques is the potential loss of EEG data. Berg and Scherg (1994) showed that as the EOG signal contains contamination by EEG, EOG subtraction can thus result in a considerable distortion of relevant brain signals. Despite this, regression and adaptive filtering remain common artefact removal techniques in many applications (Sahul et al., 1995; He et al., 2004; Romero et al., 2009; Molla et al., 2012). A further drawback stems from the fact that most linear regression and adaptive filtering techniques require the use of additional physiological reference signals to remove artefacts. In the absence of these signals, or where those signals themselves become corrupted with artefact, regression and adaptive filtering techniques break down. A potential solution to this problem was proposed by Jiang et al. (2007) who used the wavelet transform in addition to adaptive thresholding to remove ECG artefacts from

EEG, without use of an ECG signal.

Blind source separation

In Section 2.6, a number of manual and semi-automatic methods of EEG artefact removal using blind source separation were discussed. However, if an artefact removal stage is to be incorporated with an automatic neurological event detection system, the artefact removal stage must too be automated. Several fully automated BSS artefact removal techniques have been presented in the literature, and offer potential for pre-processing the EEG in automated neurological event detection systems.

One approach to automating artefact removal using BSS has been to take advantage of additional physiological signals that measure the artefact in question. In this manner, Jung et al. (2000) combined ICA and linear regression with the EOG signal to remove ocular artefacts. Similarly, Klados et al. (2011) introduced REG-ICA which used regression of the EOG signal and the independent components in order to remove ocular artefacts. Park et al. (2003) used ICA to automatically remove ocular artefacts by removing components that were highly correlated to the EOG reference signals. Joyce et al. (2004) proposed a similar technique to remove ocular artefacts using ICA, the EOG signal and some frequency spectrum thresholding. James and Gibson (2003) proposed temporally constrained ICA as a technique that incorporates the reference signal into the ICA algorithm to remove ocular and ECG artefacts from the EEG. Similarly, Devuyst et al. (2008) used a modified version of ICA to implement an automatic ensemble average subtraction of the ECG signal from independent components to remove ECG artefacts. De Vos et al. (2011) proposed a combination of filtering and correlation with the relevant polygraphy signal to remove cardiac and respiration artefacts from neonatal EEG.

A disadvantage of many automated ICA artefact removal techniques is the need for additional physiological signals. In many settings the availability of such signals is not possible. Delsanto et al. (2003) and Mognon et al. (2011) implemented automatic recognition of ocular artefacts extracted by ICA using a combination of spectral features, spatial topography, and time-domain signal morphology, without use of the EOG reference signals. Nazarpour et al. (2008) used *a priori* information regarding the spatial distribution of the ocular artefact to aid in its identification, also without use of the EOG reference signals. Similarly, Zhou and Gotman (2009) proposed an automated ICA method to remove ocular artefacts without the EOG signal, based on a dipole model of the eye. Faul et al.

(2005b) introduced an ICA-based method where artefacts were indirectly removed when ICs were ranked based on features which were known to describe neonatal seizure. Zhou and Gotman (2004) used a combination of ICA and wavelet transform to automatically remove muscle artefacts without the use of additional physiological signals.

In the above methods, heuristics and thresholds were used to identify artefactual independent components; however, machine learning techniques outperform simple thresholding when the feature in question does not clearly separate between classes (i.e. artefact and non-artefact ICs). That is to say, machine learning algorithms perform separation based on several features at once, where greater separability may exist. For typical EEG artefacts a single feature is not enough to provide sufficiently accurate separation between normal, seizure and artefact EEG. Shoker et al. (2005) proposed a fusion of blind source separation and support vector machines to remove ocular artefacts from the EEG. Halder et al. (2007) proposed a similar method for a BCI application. LeVan et al. (2006) combined blind source separation in the form of ICA with a Bayesian classifier to effectively detect artefacts in ictal EEG. Shao et al. (2009) paired ICA with a number of different classifiers to remove ocular and ECG artefacts. More recently, Nolan et al. (2010) and Winkler et al. (2011) introduced methods of automatically removing EEG artefacts for an event-related potential and BCI applications, using ICA and statistical thresholding of features which are known to describe artefact. These methods hold considerable potential for artefact processing in automated EEG diagnostic systems; however, they rely on accurately annotated ICA-decomposed data, something which is often unavailable.

Principal component analysis (PCA) is another statistical technique that has been used for blind source separation of the EEG into uncorrelated estimated source components. Lins et al. (1993) and Lagerlund et al. (1997) showed that using PCA, ocular artefact components can be identified and removed, especially if their amplitude is high. More recently, Schachinger et al. (2007) used PCA to decompose the signal into sub-bands and applied an adaptive filter to reduce high-amplitude artefacts. However, PCA suffers from a number of restrictions in removing EEG artefacts. Lamothe and Stroink (1991) showed that there does not exist a one-to-one correspondence between principal components and individual cerebral sources, making it generally inappropriate to ascribe physiological significance directly to the individual components. Furthermore, Lagerlund et al. (1997) showed that PCA cannot completely separate eye-movement artefacts from the EEG signal when they have comparable amplitudes. Note that PCA refers here to its direct use in blind source separation; PCA is also often used for dimensionality reduction, transforming data into

a lower dimension representation based upon the directions of maximal variance in the data. Within the scope of EEG processing, PCA is often used as a preprocessing step to ICA (to select only those components with the most variance for ICA). Alternatively, PCA is often used as a feature reduction technique before machine learning classification (to select those features with the most variance).

Several other artefact removal techniques have been applied to EEG data with varying success. Gao et al. (2010) showed that canonical correlation analysis is more suitable than ICA at removing EMG artefact from the EEG and implemented an automated artefact removal system on simulated data to do so. More recently Zhang et al. (2012) demonstrated the use of CCA and the EOG channel to automatically remove eye blink artefacts from the EEG.

The artefact detection and removal techniques outlined here illustrate the wide range of approaches that have been applied in an attempt to mitigate the effects of artefact. However, for the most part these approaches have not been demonstrated for their ability to deal with artefacts in automated EEG diagnostic systems. Indeed in many of the outlined methods, shortcomings in the nature of the data and the manner in which the algorithm is evaluated, mean that the reported results may be significantly over-optimistic in predicting how the artefact processing methods would perform as part of an automated neurological event detection system. This issue will be discussed in more detail in Section 2.9. Before this is done, the following sections will introduce the background of the two neurological event detection tasks that are investigated in this thesis: automated epileptiform activity detection and automated neonatal seizure detection. A brief overview of the published literature on artefact processing methods in these paradigms is then provided.

2.8 EEG artefacts in automated EEG diagnostic systems

In this thesis, the role of artefact detection and removal is evaluated in the context of two automated medical EEG diagnostic systems: epileptiform activity detection and neonatal seizure detection. While the two classification tasks are similar in nature, there are a number of significant differences between them; most notably, epileptiform activity detection requires the detection of very short, often infrequent neurological events where it is imperative to detect all possible occurrences, whereas neonatal seizure detection requires the detection of longer seizure activity, with the emphasis on very low numbers of false

detections. Consequently, different artefact processing approaches may be better suited to each neurological event detection system. The following paragraphs outline the problem that these systems aim to solve, as well as providing a literature review of the published solutions to date. In both cases, the role of automated artefact processing is also reviewed.

2.8.1 Epileptiform activity and epileptic seizure detection

Epilepsy

Epilepsy is one of the most common neurological disorders in humans, affecting almost 50 million people worldwide (Sanei and Chambers, 2007). It is defined as an underlying tendency of the brain to produce sudden bursts of abnormal electrical activity that disrupt other brain functions. Well-known causes of epilepsy may include: genetic disorders, traumatic brain injury, metabolic disturbances, alcohol or drug abuse, brain tumour, stroke, infection, and cortical malformations (Jokeit and Schacher, 2004).

The electroencephalogram is the primary tool used in the diagnosis of the disease. Diagnosis is based on the presence of epileptic seizure activity (Figure 6.1)shows an example of general idiopathic seizure in an adult patient), which can last from several seconds to an hour, and also based on the presence of shorter duration interictal, epileptiform activity (Figure 6.2), which is an indicator of susceptibility to epileptic seizure (Chatrian et al., 1964). An epileptic seizure is a disturbance characterized by changes in neuronal electrochemical activity that results in abnormal synchronous discharges in a large cell population, giving rise to clinical symptoms and signs. Patients experience varied symptoms during seizures depending on the location and extent of the affected brain area. Symptoms may include involuntary clonic movements, an altered state of awareness, convulsions, or impairment of consciousness, unusual or repetitive behaviours, or odd sensations. Seizures can be classified in several ways based upon the site and extent of the brain that is affected, clinical symptoms, EEG pattern, and etiology (Holmes, 1997). Partial seizures are limited in extent; generalized seizures typically affect the entire brain and impair consciousness. Epileptiform activity by contrast are shorter duration (usually less than a second) abnormal waveforms with no clinical signs (Binnie and Stefan, 1999).

Following a referral from a neurophysiologist, the initial test for epilepsy is known as a routine EEG, which lasts 20-40 minutes and looks for signs of abnormal waveforms in the recording (including some or all of the following features: slow waves, spikes and sharp

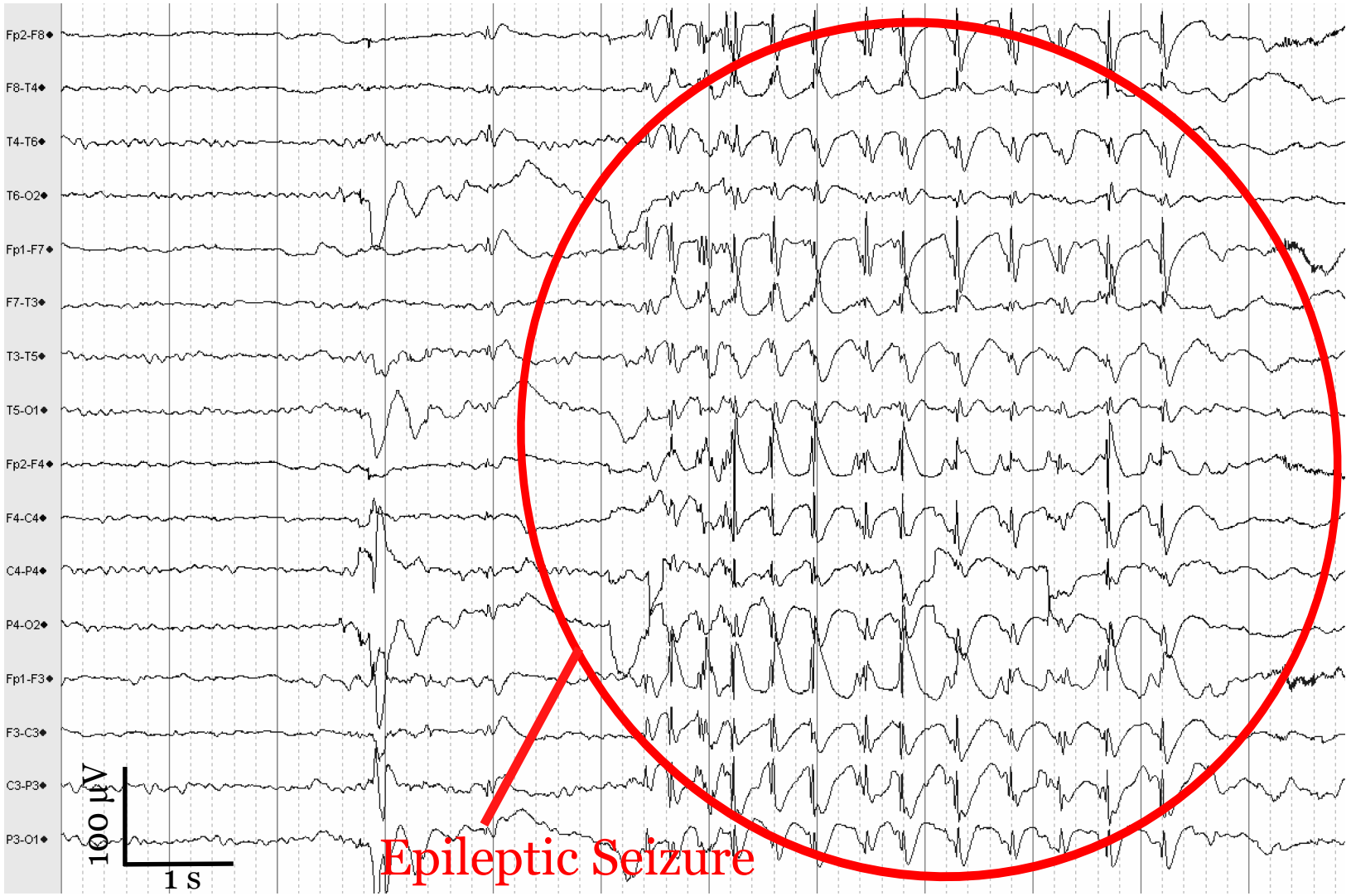


Figure 2.11: EEG showing an example of a generalised idiopathic epileptic seizure recorded from an adult patient at CUH. The data was recorded using a Viasys Nicolet EEG machine with a sampling frequency of 256 Hz.

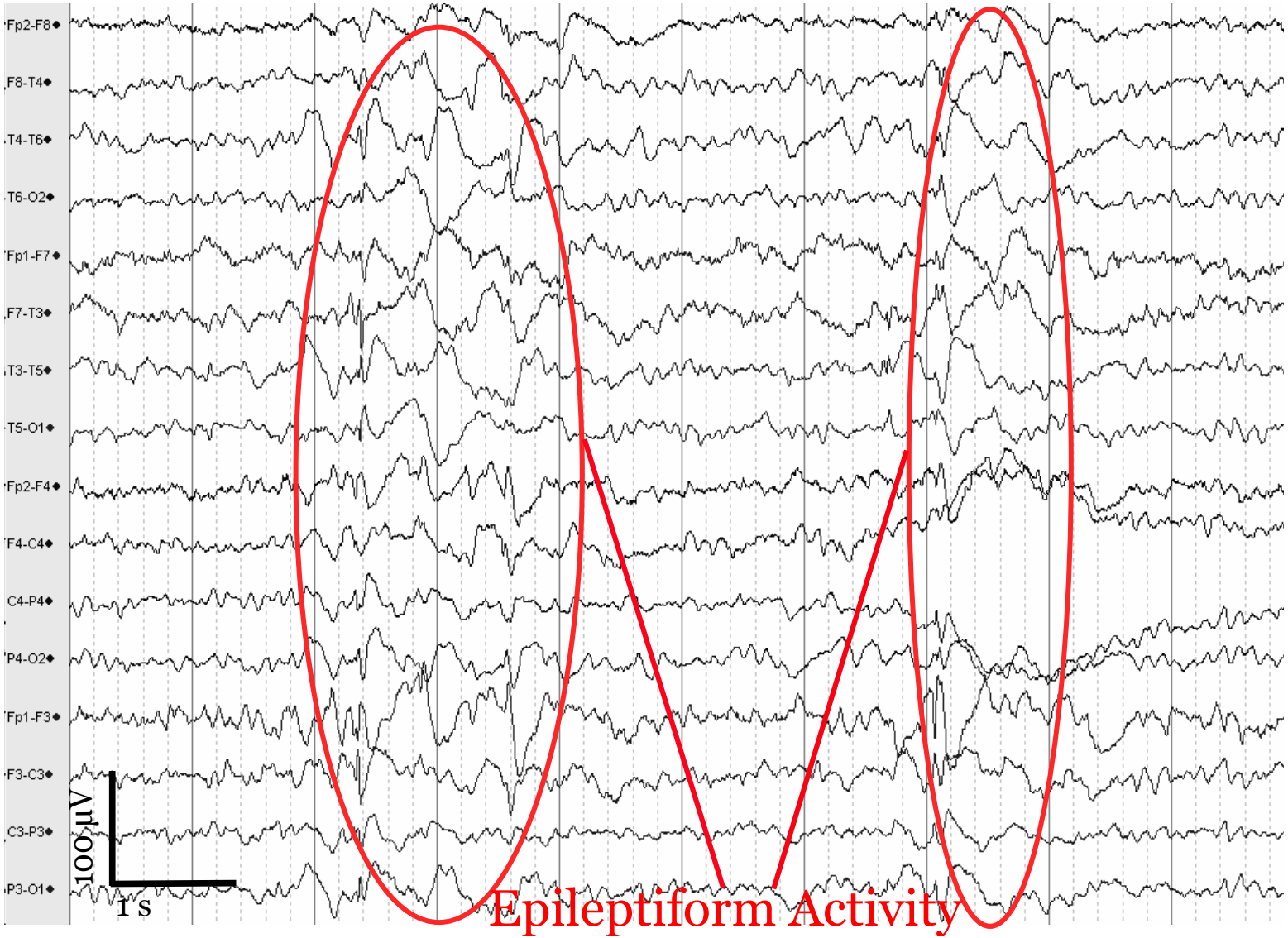


Figure 2.12: EEG sample showing two instances of epileptiform activity recorded from an adolescent patient at CUH. The data was recorded using a Viasys Nicolet EEG machine with a sampling frequency of 256 Hz.

waves, spike-wave complexes, and polyspikes). If, at the end of the test, no such activity has been recorded and epilepsy is still suspected, the patient may be asked to return to the hospital for a longer, continuous EEG (usually lasting 24-72 hours); this recording may be ambulatory and there is considerable evidence to suggest that it is clinically desirable for it to be so (Seneviratne et al., 2013). The recorded signal is then visually inspected by a neurophysiologist for abnormalities.

Automated epileptic seizure and epileptiform activity detection

Due to the costly, time-consuming nature of the continuous EEG, it has been proposed that an EEG system, with the capability to automatically detect epileptic seizure activity as well as the shorter duration epileptiform activity, would be useful for clinical and domestic monitoring of patients (Waterhouse, 2003; Gotman and Gloor, 1976; Indiradevi et al., 2008). Such a system would help the clinician pinpoint the exact location of abnormalities in the recording, thus vastly reducing post-recording analysis time.

Automated epileptiform detection systems can be traced back to the half-wave decomposition method proposed by Gotman and Gloor (1976). Comprehensive reviews of epileptic seizure and epileptiform detection (sometimes referred to as spike detection in the literature) algorithms have been compiled by Wilson and Emerson (2002), Casson et al. (2009), Halford (2009) and Song (2011). Highlights include: the wavelet decomposition method proposed by Indiradevi et al. (2008), the neural network seizure detection system outlined by Bao et al. (2008) and the ICA preprocessed method introduced by De Lucia et al. (2008). Recent work by Kelleher et al. (2010) proposed an SVM-based classifier as effective for epileptiform activity and epileptic seizure detection.

2.8.2 EEG artefacts in epileptiform activity detection systems

Artefacts are widespread in most EEG recordings, and result from the wide array of causes outlined in Section 2.3. In addition to routine causes of EEG artefacts, many epileptic seizures are often accompanied by involuntary clinical symptoms that introduce a significant amount of artefact (Gotman et al., 1999). These artefacts can interfere with seizure interpretation, especially if they are present at the time of the seizure onset. Artefacts are a significant obstacle in the under-performance of many automated epilepsy diagnosis algorithms (Wilson and Emerson, 2002). A number of papers in the literature propose

methods for removing artefacts in epileptic EEG. A number of manual and semi-automated artefact removal techniques have been described in the literature (Section 2.6). Iriarte et al. (2003), Urrestarazu et al. (2004) and Liu et al. (2004) successfully used ICA-based semi-automatic methods of removing artefacts from EEG recordings of patients undergoing epileptic seizure without introducing obvious distortion to the seizure morphology. Ille et al. (2002) used PCA to a similar effect. De Clercq et al. (2006) and Vergult et al. (2007) proposed semi-automated CCA methods to remove EEG muscle artefact for patients with refractory partial epilepsy.

A number of papers published on automated epileptic seizure and epileptiform activity detection include automated artefact detection and removal. In their half-wave decomposition method of interictal epileptiform activity detection, Gotman and Gloor (1976) introduced some rule-based thresholding techniques to reject sections of EEG contaminated with EMG and ocular artefacts. The techniques outlined in the paper were tailored to the epileptiform detection system to reject specific cases of false positive detection of spike and sharp wave activity. These methods were later updated for epileptic seizure detection (Saab and Gotman, 2005). Kuhlmann et al. (2009) employed these methods with slight modifications in their paper on seizure detection using seizure probability estimation. In their rule-based epileptiform detection system, Dingle et al. (1993) included threshold-based rules tailored to exclude sections of artefactual EEG. A similar rule-based methodology was employed by Ramabhadran et al. (1999) in their automated epileptogenic focus localisation system. This approach of dealing only with those artefacts that are responsible for false detections is particularly suited to reducing the number of false detections in epileptiform activity detection as it is important to continue to maintain correct detections of epileptiform events; however, the rule-based approaches listed do not take advantage of some of the most powerful aspects of machine learning. An alternative approach to tackling problematic artefacts in epileptiform activity detection systems is to use blind source separation at the pre-processing stage. In this manner, Hesse and James (2007) and De Lucia et al. (2008) used ICA to automatically detect epileptiform activity in epileptic patients. By applying this approach, a number of EEG artefacts were implicitly excluded, i.e. epileptiform detection was performed on ictal independent components. However, these systems were either demonstrated for very short EEG segments (Hesse and James, 2007) or performed significantly below state-of-the-art (De Lucia et al., 2008).

2.8.3 Neonatal seizure detection

Neonatal seizure

The EEG can be used to monitor sick newborn patients who are admitted to the neonatal intensive care unit (NICU) and are at risk of developing neonatal seizures (Murray et al., 2008). These neonatal seizures are clinically defined as paroxysmal alterations in neurological function, i.e. an alteration in behavioural, motor and/or autonomic function (Volpe, 1989). Hypoxic ischemic encephalopathy (HIE) is the most common neonatal seizure etiology (i.e. cause); induced by impaired gas exchange that results in inadequate oxygen supply to the brain (hypoxia) and neuronal damage (ischemia) (Volpe, 2008; Pin et al., 2009). Other common etiologies are central nervous system infection, cerebral infarctions and metabolic abnormalities (Evans and Levene, 1998; Ronen et al., 1999).

Where visible on the EEG, these seizures manifest as repetitive patterns with minimum duration defined as 10 seconds (Clancy, 2006). A major confounding factor in the detection of neonatal seizures is the fact that seizures may not be clinically observable, especially if the patient is prescribed with an anti-epileptic drug (Clancy et al., 1988; Bye et al., 1997). The EEG is therefore considered the gold standard in neonatal seizure diagnosis. However, the diagnosis of electrographic seizure is non-trivial. The appearance of the discharge can transform in frequency, amplitude and morphology over time (Mizrahi and Clancy, 2000). Electrographic seizures are predominantly unifocal but multifocal seizures may also occur in different brain regions and fire simultaneously and asynchronously. Both unifocal and multifocal seizure discharges can spread from one location via abrupt change or by gradual widening.

Automated neonatal seizure detection

Strong parallels exist between automated neonatal seizure detection and automated epileptic seizure detection. These research areas can be traced back to the automated method of recognising inter-ictal epileptic activity in EEG by Gotman and Gloor (1976). Automated neonatal seizure detection first appeared in the literature with an approach by Liu et al. (1992) to quantify the periodicity of the EEG during seizure via autocorrelation analysis. Gotman et al. (1997) introduced a threshold-based neonatal seizure detection system, where the majority of features were extracted from the frequency domain. Roessgen et al. (1998) introduced a model-based approach to seizure detection. Celka and

Colditz (2002) proposed a seizure detector based on complexity analysis of the EEG. A number of systems have been developed using wavelet analysis and classification to detect neonatal seizure (Nagasubramanian et al., 1997; Zarjam et al., 2003; Hassanpour et al., 2004a,b). Faul et al. proposed several techniques for automated neonatal seizure detection, including an ICA pre-processing stage (Faul et al., 2005b), complexity features (Faul et al., 2005a) and a neural network classification stage (Faul, 2007). A number of classifier-based methods for neonatal seizure detection have shown promising results. Karayiannis et al. (2006) proposed a rule-based architecture incorporating a neural network classifier, Aarabi et al. (2007) proposed a multi-class neural network with 6 non-seizure and 2 seizure states. Greene et al. (2008) investigated the performance of a number of 2-class classifier architectures. More recently, Thomas et al. (2010) and Temko et al. (2011b) proposed a neonatal seizure detection system based on Gaussian mixture models and support vector machines respectively.

Several multimodal seizure detection systems have been outlined in the literature. A video-based method to quantify motor activity during clinical seizures was proposed by Karayiannis et al. (2001). Data fusion methods combining seizure information from EEG and ECG were developed by Greene et al. (2007) and Malarvili and Mesbah (2008).

At present, the relatively poor performance of automated seizure detectors has prevented their advance into widespread clinical use. In reviewing what features have prevented this transition from research literature to clinical implementation, Thomas (2011) has pointed out that neonatal seizure is composed of a large set of diverse patterns emanating from background activity, seizure events and artefacts. Additionally, EEG characteristics are often considerably different between patients, and are not stationary, i.e. they evolve over time, particularly in the immediate weeks after birth. However, Thomas (2011) found that the largest contributor to false detections was that of artefact, accounting for 43 % of false positive detections of neonatal seizure. In particular loose electrode or electrode detachment were problematic, leading to significant 50 Hz line noise and subsequent motion artefact as the electrode was free to move. Respiration and movement artefacts were also highly prominent and problematic. In a detailed clinical study of the SVM classifier performance on a neonatal cohort collected at University College London Hospitals, Mathieson (2012) found that contamination of the neonatal EEG by non-cerebral electrical artefacts is the single largest cause of misclassification in the automated systems developed by Faul (2007), Thomas (2011) and Temko et al. (2011b). Similarly, in the evaluation of their neonatal seizure detection algorithm Gotman et al. (1997) found that

artefacts were responsible for the largest fraction of false detections.

2.8.4 EEG artefacts in neonatal seizure detection

Many automated neonatal seizure detection algorithms employ artefact detection and removal techniques. There are clear parallels between these artefact processing techniques and those designed for automated epileptic seizure and epileptiform activity detection systems. Witte et al. (1987) implemented a regression method for ECG and EOG artefacts in neonatal EEG. Gotman et al. (1997) incorporated the rule-based thresholding method from Gotman and Gloor (1976) into a neonatal seizure detection system. More advanced rule-based artefact rejection schemes were subsequently implemented by Glover et al. (2002) and Mitra et al. (2009). Aarabi et al. (2007) used an array of artefact techniques to remove individual artefact types. Template matching was used to remove ocular and ECG artefacts; thresholding in the time domain was used to remove electrode movement and pop, and thresholding in the frequency domain was used to remove EMG artefact. However, the effect that each of these artefact processing rules produced on the epileptiform detection systems was not quantified. Faul et al. (2005b), used ICA to indirectly remove artefacts by performing feature ranking and keeping only those ICs that most resembled neonatal seizure.

Multimodal artefact rejection was used by Boashash et al. (1999) by combining thresholding methods from the EEG, EOG, ECG and video. Karayiannis et al. (2006) and Mitra et al. (2009) used similar approaches by including a post-processing stage where seizure decisions were rejected if the corresponding section of EEG was highly correlated to any of the ECG, EMG or respiration signals. De Vos et al. (2011) proposed an automated ICA method to remove ECG, pulse and respiration artefacts. In this multimodal technique, the independent components were compared to filtered polygraphy signals and removed if the correlation exceeded a pre-defined threshold. These ICA-based multimodal correlation methods proved effective in the system of De Vos et al. (2011); however, the reliance on the presence of a respiration trace is problematic, as they are not available in many NICUs.

2.9 Datasets and metrics

The primary goal of the work detailed in this thesis is to advance the development of artefact processing in automated EEG diagnostic systems. In doing so, it is hoped that the performance of these neurological event detection system will improve. To accomplish this, data is necessary. Moreover, it is imperative that the data is representative of the data that will be encountered in real-world scenarios.

While the published literature in artefact detection and removal have undoubtedly provided a wealth of information and ideas regarding the processing of artefacts in EEG, there are a number of methodological shortcomings that are visible across many of the papers. Most notably, biases and assumptions regarding the datasets upon which the algorithms are developed and evaluated are often ignored or downplayed. Bias can enter the datasets from a number of directions. Firstly, is the data representative of that which will be encountered in the real-world? The majority of papers published in the literature that deal with artefacts in EEG, do so on simulated, pruned, or otherwise unrepresentative EEG data. Similarly, many papers evaluate the performance of algorithms on data whose duration is of the order of seconds, rather than minutes, hours or even days. Furthermore, there is often no discussion as to whether the included data was randomly chosen or selected as it showed favourable performance with the artefact processing technique in question. This thesis is founded upon the belief that in order to deal effectively with the very real problem of EEG artefacts, real data, representative of that encountered in noisy recording environments, should be used to train, and more importantly test the signal processing and machine learning techniques used to identify and/or remove the unwanted artefacts. Where this is impossible, the limitations of the conclusions and generalisability of the algorithm should be clearly stated.

The work outlined in this thesis follows three broad threads: (i) detecting head-movement artefacts so that automated neurological event detection systems in ambulatory EEG can be developed, (ii) processing of ocular and movement artefacts in an automated epileptiform activity detection system to reduce false detections and (iii) removing respiration artefacts in order to improve a state-of-the-art neonatal seizure detection algorithm. Separate, specific datasets are used for each of these tasks; in each case, considerable effort was made to ensure that these datasets would be as representative as possible of those encountered in real-world, clinical and ambulatory environments.

The first data set is a non-clinical, ambulatory EEG data set collected at UCC in 2009 using the Emotiv EPOC headset, and outlined in Section 2.9.1. The second dataset used in this thesis, outlined in Section 2.9.2, is a clinical dataset collected from epileptic patients in the neurophysiology department of Cork University Hospital (CUH). The final dataset consists of clinical data collected from neonates at the neonatal intensive care units at Cork University Hospital and University College London Hospitals, and is described in Section 2.9.3.

2.9.1 Head-movement artefacts in adult EEG

Prior to the work carried out in the thesis, there existed no attempt in the literature to deal with head-movement artefacts in particular or indeed artefacts in an ambulatory environment in general. In order to investigate and experiment with potential methods for EEG artefact detection and removal, a non-clinical dataset was collected using the Emotiv EPOC headset (Emotiv EPOC headset). The 14-channel Emotiv EPOC is a commercially available EEG headset recently released for the development of BCI technologies. The headset records EEG from 14 channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4), sampled at 128 Hz, with a referential montage employed; the reference electrodes (P3 and P4) are taken from behind the ears. Figure 2.13 compares EEG recorded using the EPOC and that recorded using a Viasys Nicolet EEG machine; it can be seen that the EPOC provides EEG of comparable quality to clinical EEG machines. A key factor in the choice of the Emotiv EPOC headset was the inclusion of gyroscopes in the device. Gyroscopes are devices that measure angular rotation, and consequently offer a means of accurately detecting head movements. Figure 2.14 shows an section of head movement with one EEG channel (blue) and two gyroscope signals (black and red) displayed. As gyroscopes or accelerometers are not included on any commercially available, clinical EEG machines, and ethics approval for their deployment in a clinical environment may prove difficult, the Emotiv EPOC offered a suitable, inexpensive alternative. Gyroscope signals from the 2-axis gyroscope located at the rear of the headset were sampled at 128 Hz. For both EEG and gyroscope signals, the amplitude resolution was $0.51\mu V$ per bit. The head-movement artefact dataset was created with the purpose of accurately capturing the types of head-movement artefacts that will be likely to be encountered in a real-world ambulatory setting. Accordingly, two alternatives were available for recording the data: (i) record the data naturally by allowing the participant to go about their daily life as normal and examine the data afterwards to annotate for head-movement artefact

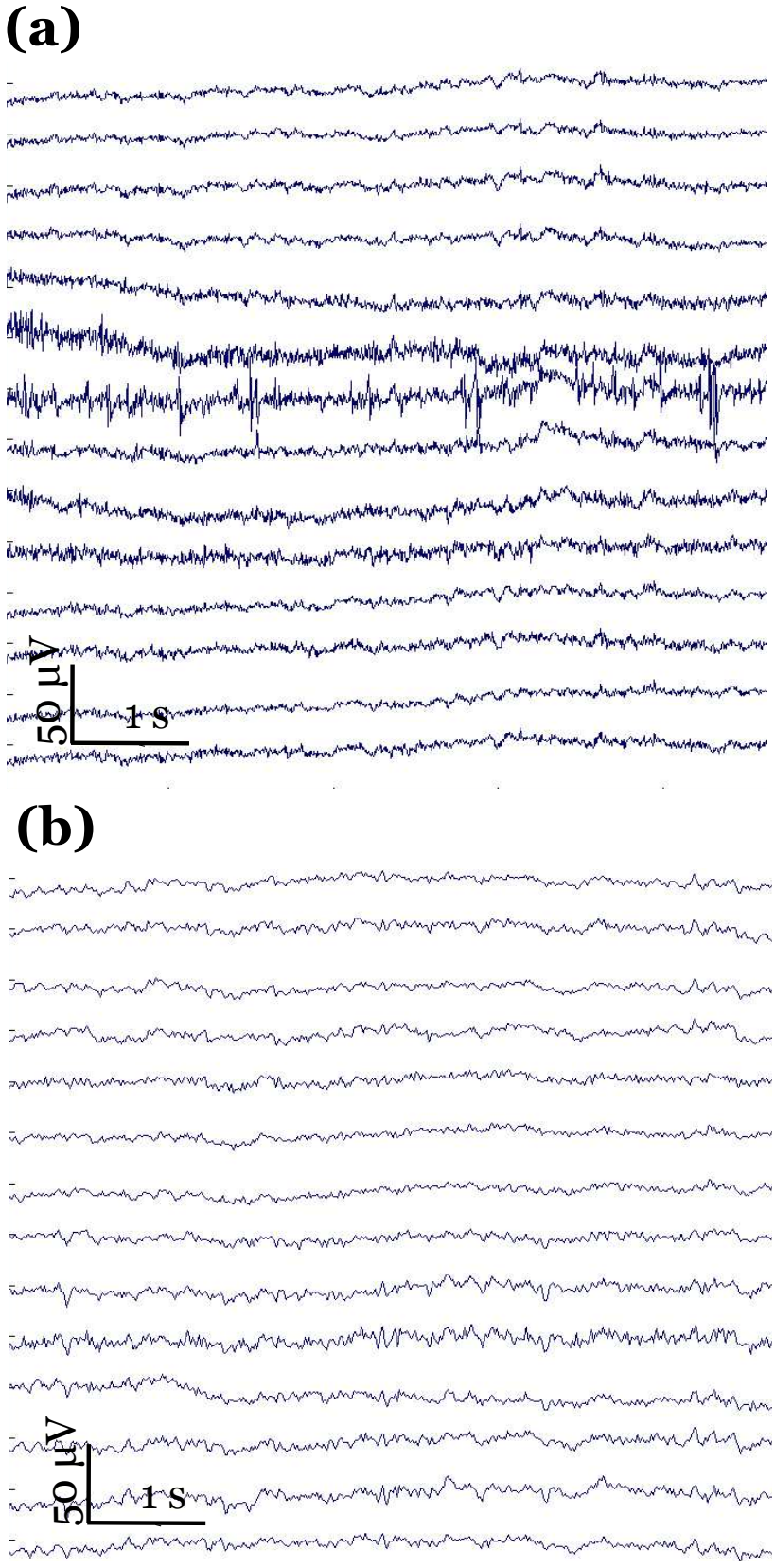


Figure 2.13: Comparison of EEG recorded using Emotiv EPOC and Viasys Nicolet One EEG machines. Approximately 5 seconds of normal background EEG are displayed for each system. (a) Normal background EEG recorded using Viasys Nicolet EEG machine at a sampling frequency of 250 Hz. (b) Normal background EEG recorded using the Emotiv EPOC EEG device at a sampling frequency of 128 Hz.

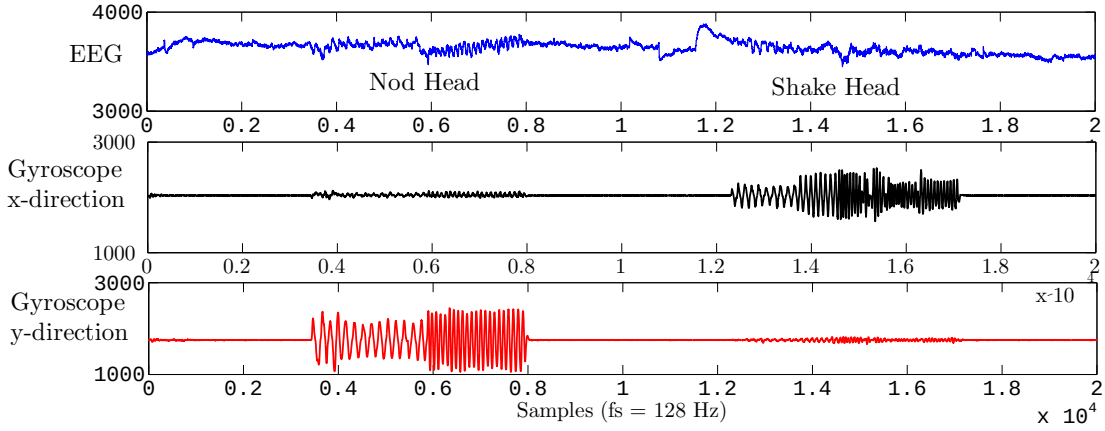


Figure 2.14: EEG channel F8 (blue) and gyroscope x-direction (black) and y-direction (red) signals for artefacts caused by nodding and shaking head.

or (ii) create a head-movement artefact protocol that described typical head-movements and instruct participants to carry out the movements at defined times. Both options were initially evaluated in terms of accuracy and difficulty and it was found that the primary obstacle in using the data, was that of annotation, i.e. how are artefact and normal EEG determined. With data collection method (i), it was extremely time consuming to examine the EEG and annotate for artefacts after the fact. Importantly, there was no discernible distinction between the types of artefact created in data collection (ii) as to those collected naturally in (i). Thus, a head-movement artefact generation protocol was used in the creation of an artefact database so as to aid in the annotation of the data.

An artefact generation protocol was drawn up which instructed the participants to perform repetitions of each of the following movements: shake head, nod, roll head, clench jaw, raise and lower eyebrows. Between repetitions participants were asked to remain still in order to generate reference EEG. Particular focus was placed on movement artefacts that have been observed to occur more regularly in an ambulatory EEG system. This artefact generation protocol is described in more detail in Table 2.1. The pace and direction of head movements were varied where appropriate in order to avoid excessively repetitive, periodic artefacts that may be unlikely to occur in a natural ambulatory environment. In total, this data comprises over 30 minutes of head movement data, collected from 7 male healthy adults (23- 50 years, mean age 30). Informed consent was obtained from all participants, none of whom had a history of neurological or psychiatric disorders and none were on chronic medication.

The EEG artefact data was annotated as artefact where visually noticeable deflection in

the EEG was observed at the times that participants performed movements. Annotation was independently confirmed by a specialist clinical physiologist in neurophysiology.

2.9.2 Ocular and movement artefacts in epileptiform EEG

The first clinical dataset used in this study consists of multi-channel EEG recordings obtained from 8 patients (9-32 years; mean age 16), each suffering from idiopathic generalised epilepsy. The data was acquired using a NicoletOne clinical EEG machine at the Department of Neurophysiology, Cork University Hospital from patients undergoing routine EEG tests. Each patient displayed several instances of either interictal epileptiform activity, epileptic seizure activity, or both. The data was sampled at 250 Hz using the 10-20 system of electrode placement and analysed using a 16-channel bipolar montage. This research work has been approved by the Clinical Research Ethics Committee of the Cork Teaching Hospitals (CREC).

A total of 42 abnormal events (consisting of single focal sharp-wave and spike and slow-wave activity) were annotated on a per-channel basis by a trained neurophysiologist. A detailed account of the data is illustrated in Table 2.2. Examples of ocular and movement artefact were also obtained from the data set. The ocular artefact data consists of 35 expert-annotated events with a total duration of 38.9 seconds, recorded at the 4 frontal EEG channels (FP2-F4, FP2-F8, FP1-F3 and FP1-F7). The movement artefact data comprises 14 expert-annotated events of total duration 79.5 seconds, taken from all 16 EEG channels.

Table 2.1: Artefact Generation Protocol.

Head-movements	*Description*	*Duration*
Shake head	Shake head from side to side, varying pace and direction	30 seconds
Remain still	Remain seated, avoiding head movements and eye blinks and movements	20 seconds
Nod head	Nod head up and down, changing pace as doing so	30 seconds
Remain still	Remain seated, avoiding head movements and eye blinks and movements	20 seconds
Roll head	Roll head in both directions, changing pace as doing so	30 seconds
Remain still	Remain seated, avoiding head movements and eye blinks and movements	20 seconds
Clench jaw	Prolonged as well intermittent clenches (mimicking chewing)	30 seconds
Remain still	Remain seated, avoiding head movements and eye blinks and movements	20 seconds
Raise and lower eyebrows	Changing pace and amplitude throughout	30 seconds
Remain still	Remain seated, avoiding head movements and eye blinks and movements	20 seconds

Table 2.2: Data characteristics for each patient record used in the development of the epileptiform activity detection system.

Patient	Record Length (s)	No. Abnormal Events	Total Event Duration (s)	Mean Event Duration (s)
1	953	2	0.8	0.4
2	1168	1	0.4	0.4
3	2736	5	3.6	0.72
4	3122	2	0.7	0.35
5	1219	13	18.2	1.4
6	1213	4	4.8	1.2
7	1221	14	29.8	2.129
8	1200	1	3.9	3.9
Mean	1604	-	-	1.312
Total	12832	42	62.2	-

2.9.3 Neonatal EEG

The neonatal EEG used in this thesis is clinical data recorded in NICUs at Cork University Maternity Hospital (CUMH) and University College Hospitals London. The dataset can be split into two main cohorts; EEG data used to develop the automated neonatal seizure detection algorithm (carried out by Faul (2007), Greene (2007), Thomas (2011) and Temko et al. (2011b)) and EEG data used to develop and evaluate the artefact removal algorithms.

Development of the automated seizure detection system

Data recorded at CUMH was used to develop, train and initially test the automated neonatal seizure detection algorithm that is utilised in this thesis. EEG was recorded from 55 babies with hypoxic ischemic encephalopathy (HIE) between 2003 and 2006; up to 72 hours of data was recorded per patient. Within the group, 17 patients underwent seizure, with all seizures that occurred over the initial 72 hour period after birth captured. These patients were full-term neonates ranging in gestational age from 39 to 42 weeks. A NicOne video EEG machine was used to record multichannel EEG at 256 Hz using the 10-20 system of electrode placement modified for neonates. In this study, 8 bipolar EEG channels were used (F4-C4, F3-C3, T4-C4, C4-CZ, CZ-C3, C3-T3, C4-O2 and C3-O1). The dataset contained over 267 hours of EEG from which a total of 705 seizure events with a mean duration of 3.89 minutes were annotated by a neonatal electroencephalographer. Per-channel annotations were necessary for classifier training; per-channel labels were obtained for 2 minutes of seizure data from each patient. Additionally, 2 minutes of artefact free non-seizure data was annotated for each patient. Using this data, the seizure detection system in its present form was developed and evaluated by Temko et al. (2011b)

Table 2.3: Data characteristics for each neonatal patient used in the development of the neonatal artefact removal algorithms.

Patient	Record length (hours)	No. of seizures	Total seizure duration (s)	Mean seizure duration (s)
1	24	20	1510	75.5
2	48	46	2963	64.41
3	24	18	1435	79.72
4	24	40	3571	89.28
Mean	30	31	-	77.23
Total	120	124	9407	-

and Thomas et al. (2011).

Data used in artefact removal techniques

Data collected at UCLH was used to develop and test methods of artefact removal at the pre-processing stage of the seizure detection algorithm described in the previous paragraphs. This data consisted of four neonatal patients; details of seizure occurrences are displayed in Table 2.3.

Each patient suffered from seizure within 72 hours of birth, and the EEG recordings of each patient contained considerable respiratory artefact. Respiration traces and ECG signals were recorded for patients 1 and 2. Patients 3 and 4 did not have a respiration trace available; ECG signals were however, recorded for these patients. As the performance of the seizure detection algorithm was the ultimate measure of the algorithms performance, annotations were not required for the respiration artefact.

2.9.4 Performance assessment

A number of metrics are used in this thesis to effectively measure and compare the performance of artefact processing algorithms. Evaluating the performance of supervised learning algorithms is reasonably straightforward, i.e. the detection algorithms are evaluated by comparing the classification output of each epoch of test data with the corresponding annotations. Additional metrics can then be constructed that aim to capture the intricacies of the classification task at hand, be it artefact detection (Chapters 3, 4 and 5), epileptiform activity detection (Chapter 6) or neonatal seizure detection (Chapter 7). These metrics are briefly outlined in this section, with additional details included in the

relevant chapters.

Detection metrics

Epoch-based metrics are used to evaluate classifier performance across a range of operating points. Each of the classification tasks described in this thesis, are formulated as binary decision problems and as such can be represented in a structure known as a contingency table. To illustrate this, consider the binary classification of EEG into two classes: artefact and non-artefact. This contingency table has four categories: true positives (TP) are epochs correctly labelled as artefact; false positives (FP) refer to epochs incorrectly labelled as artefact; true negatives (TN) correspond to correctly labelled non-artefact epochs and false negatives (FN) refer to epochs incorrectly labelled as non-artefact. The accuracy of each class is then evaluated using *Sensitivity* and *Specificity*. *Sensitivity* is defined in equation 2.12 as:

$$Sensitivity = \frac{TP}{TP + FN}. \tag{2.12}$$

Specificity is defined in equation 2.13 as:

$$Specificity = \frac{TN}{TN + FP}. \tag{2.13}$$

Receiver Operator Characteristic (ROC) curves show how sensitivity varies with specificity; with the area under the ROC curve shown to be an effective way of comparing the performance of two different classifier systems (Fawcett, 2006). The performance of the artefact detection systems outlined in Chapters 3, 4 and 5 are compared using ROC curves, and the area under the ROC curve. In this thesis, areas under the ROC curve are calculated using trapezoidal integration.

In the neurological event detection systems described in this thesis (epileptiform detection and neonatal seizure detection) it is important to evaluate the classification performance in the region of the ROC curve most relevant to the classification task. That is, by incorporating domain-specific information about the neurological event as well how the neurological event detection system will be utilised, allows more effective evaluation of system performance. For the epileptiform detection systems outlined in Section 2.8.1, it is important to detect all (or almost all) epileptiform events, even if this means large numbers of false detections; as epileptiform events are rare and often short, missing an

epileptiform epoch can mean missing the epileptiform event in its entirety. Missing entire epileptiform events is unacceptable in the context that the system will be used, namely screening long EEG recordings for indicators of susceptibility to epilepsy. That is, missing an event means the event will never be seen by the clinician reviewing the EEG, whereas false detections (within reason) can simply be discarded upon inspection. With this in mind, it is clear that the area of the ROC curve where sensitivity is high is primarily of interest. An additional metric, ROC_{sens95}, is thus used to measure the classification performance of each epileptiform activity detection classifier. ROC_{sens95} represents the area under the ROC curve above a sensitivity of 95 %, where a ROC_{sens95} of 100 % refers to perfect discrimination in this region of the ROC curve. In Figure 2.15 it can be seen that a random discrimination is highlighted by the dotted red line, and is equivalent to a ROC_{sens95} of 2.5 %. As was the case for the ROC areas reported in this thesis, the ROC_{sens95} area was calculated using trapezoidal numerical integration.

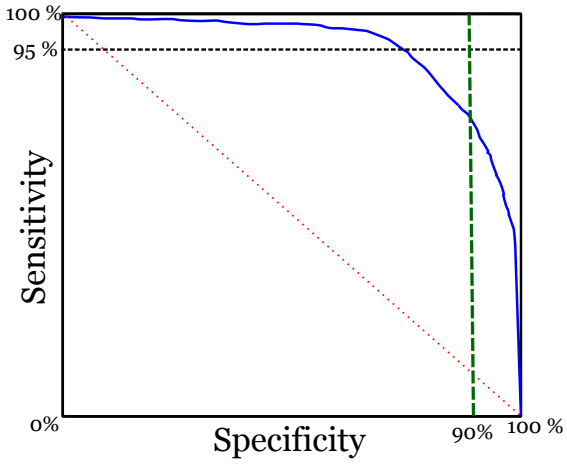


Figure 2.15: Illustration of ROC_{sens95} and ROC_{spec90}. The ROC curve is plotted in blue. The horizontal, dashed line represents the lower bound of the ROC_{sens95} area used in the epileptiform activity detection algorithm. The vertical green line represents the lower bound of the ROC_{spec90} area, used in evaluating the neonatal seizure detection algorithm. The dotted red line represents random discrimination in the classification task.

In the neonatal seizure detection systems described in Section 2.8.3, the necessity to detect all seizure epochs can be relaxed so as to facilitate a reduction in false detections. False detections may have clinical decisions (e.g. prescription of anti-epileptic drug) resting heavily upon them and so must be kept to an absolute minimum. Additionally, as seizures typically last longer than a single epoch, if one epoch in a seizure is missed, the seizure can still be detected by the classifier. The neonatal seizure detection system will thus operate in regions of the ROC curve where specificities are high; an additional metric, ROC_{spec90},

is thus used. ROC_{spec90} is defined as the area under the ROC curve, above the specificity of 90%, as illustrated in Figure 2.15.

In addition to the epoch-based metrics outlined above, there are a number of *event-based* metrics used in evaluating the neurological event detection classifiers in this thesis. *Good Detection Rate (GDR)*, is the percentage of neurological events (epileptiform activity or neonatal seizure) within which at least one epoch is correctly classified as representing the neurological activity in question, and defined in equation 2.14 as:

$$GDR = \frac{TP_{events}}{TP_{events} + FN_{events}}. \tag{2.14}$$

False Detections per hour (FD/h), another event-based metric, is used to assess the false detections of the classifier; FD/h is defined as the number of predicted seizure events in 1 hour that have no overlap with actual reference seizures (Temko et al., 2011c). A final epoch-based metric, *False detection rate (FDR)* is also used in this thesis for evaluating the performance of the epileptiform activity detection algorithm in Chapter 6. However, as the reasons for using this metric are closely linked to the details of the classifier details, a more thorough description of FDR is discussed in Chapter 6.

Artefact removal metrics

Artefact removal algorithms have two primary, and often overlapping goals; to remove artefacts from the EEG while leaving the cerebral portion of the EEG intact, so as to (i) improve the performance of a neurological event detection algorithm and/or (ii) "clean" the EEG for ease of interpretation by a clinician. Accurately evaluating the performance of artefact removal algorithms presents a less straightforward challenge than that of evaluating detection algorithms. That is to say, quantifying the extent of artefacts removed, and the corresponding degree of preserving the cerebral portion of the EEG, is difficult without prior knowledge as to what a "cleaned" EEG signal should look like. Conversely, the performance of artefact removal algorithms that use simulated EEG can be easily quantified, as the original, clean EEG is artificially mixed with corrupting artefacts. Thus, metrics that quantify the likeness of the original, clean EEG with the artefact-processed, "cleaned" EEG can be used. These metrics include the correlation, mean-square error and signal-to-noise ratio amongst others. It is presumed that this is a primary reason that many artefact removal publications are evaluated on simulated data. However, while these metrics are numerically precise, this precision does not necessarily pertain directly

to real-world artefact removal. The data recording method, the data selection process, and whether the background and artefact EEG was real or modelled (as in many publications) all introduce considerable biases into the experiment. The method in which the background and artefact EEG are mixed also introduce significant biases into the experiment, most notably when evaluating the performance of blind signal separation techniques which in turn are predicated upon assumptions about the nature of mixing of the underlying sources. These arguments coupled with those outlined in Section 2.9.1 explain why simulated artefact data was not used in this thesis.

To guide the choice of evaluation metric, it is perhaps worth re-stating that the primary research goal in this thesis is to develop artefact processing techniques that improve the performance of state-of-the-art automated EEG medical diagnostic systems. The principal measure of artefact removal performance is thus whether or not the underlying neurological event detection system improves or not. Therefore, it is this measure (and the classification metrics associated with it) that is used to evaluate the performance of respiration artefact removal for neonatal EEG (Chapter 7).

2.10 Summary

This chapter outlined in detail the pervasive issue of artefact contamination in EEG recordings. These artefacts interfere with visual inspection of the EEG and often significantly hinder the performance of automated neurological event detection systems. The sheer volume of papers published that propose methods aimed at alleviating this problem gives an indication of the necessity for automated systems to detect and remove these artefacts. A comprehensive review of the merits and failings of these algorithms, however, showed that it is also clear that despite the efforts made, current state-of-the-art in automated artefact detection and removal is not sufficient.

With EEG considered the gold standard for many applications where spatio-temporal information regarding the functioning of the brain is required, it is apparent that automated neurological event detection systems are desirable. Two such areas in the medical domain are the fields of epilepsy and neonatal seizure diagnosis. This chapter reviewed current methods for automated neurological event detection systems in epilepsy and neonatal seizure diagnosis; it is evident from this literature review that these systems have yet to attain the requisite performance necessary for widespread clinical deployment. EEG

artefacts have been widely reported as a leading obstacle in the achievement of this goal. Accordingly, this chapter outlined a framework for the biomedical signal processing and machine learning methods that are used in this work to further artefact detection and removal processing. The following chapters will detail novel methods that were developed in this thesis to detect and remove these problematic artefacts. This work will follow three broad threads: (i) detection of head-movement artefacts in ambulatory EEG, (ii) artefact detection and fusion in a clinical, epileptiform activity detection system and (iii) artefact removal in a state-of-the-art, neonatal seizure detection system.

Chapter 3

Detecting artefacts arising from head-movements

3.1 Introduction

When the head moves during an EEG recording, a wide range of electrical signals arising from sources external to the brain may contaminate the EEG signal. Contamination typically takes the form of some combination of EMG, ocular, electrode pop and electrode movement artefacts, and obscures the EEG trace, making it difficult to interpret the signal by the researcher or clinician, especially for inexperienced readers. In automated neurological event detection systems such as epilepsy detection or brain computer interface applications, artefacts introduced by head-movements frequently lead to misclassification by the pattern recognition system. This problem becomes more pronounced in ambulatory environments where the movements of the patient or user are unrestricted and where head-movements are more frequent. To date, ambulatory EEG has received little or no attention in the literature, except to point out the problem of artefact and the usefulness of ambulatory monitoring (Waterhouse, 2003). In this chapter, the feasibility of accurately detecting these head-movement artefacts using statistical pattern recognition techniques is investigated.

A journal paper (O'Regan et al., 2013a) and a peer-reviewed international conference publication (O'Regan et al., 2010a) have arisen from the work carried out in this chapter.

3.2 The problem of head-movement artefacts in ambulatory EEG

Artefacts arising from head-movements often prove troublesome both for EEG interpretation by a clinician, and also in the deployment of automatic neurological event detection systems, such as epileptic seizure detection or brain-state classification. By obscuring the EEG during epileptic seizures, artefacts can interfere with the clinicians' interpretation of the recorded seizures, often making it difficult to identify and localise the ictal onset and offset (Delorme and Makeig, 2004). In automatic seizure detection, the presence of artefacts may lead to falsely interpreting a section of artefact EEG as seizure. In many clinical EEG trials, contamination by artefacts is minimized by controlling the test situation to limit movement. In an ambulatory setting this is unrealistic as by definition ambulatory refers to an environment where the patient can walk and move about. Similarly, in cases such as diagnosing epilepsy, restricting movements may even be undesirable as it may exclude the presentation of events that occur during everyday life and that trigger epileptic seizure in the patient.

3.2.1 Generalised head-movement artefact detection: is it possible?

Head movements can introduce a wide range of non-cerebral electrical activity into the EEG. Typically these movements result in contamination in the form of some combination of muscle (EMG), electrode pop, electrode movement and ocular artefacts. As discussed in Chapter 2, these component artefact signals display a wide range of characteristics. Muscle artefacts are predominantly high frequency signals, and can range from low to high amplitude (van de Velde et al., 1999; Goncharova et al., 2003). Electrode pop is typically accompanied by fast, high amplitude spikes (Barlow, 1986). Electrode movement most commonly results in slow-wave baseline drifts, but can sometimes result in apparent oscillation in the EEG. Ocular artefacts, introduced due to relative movement between the eye and the electrode, usually result in high amplitude deflections in the EEG (Berg and Scherg, 1991; Croft and Barry, 2000; Gasser et al., 1992).

While these component artefact signals exhibit diverse temporal, frequency and structural characteristics, they are significantly different from those of normal EEG activity. It is thus proposed that the component artefact signals (EMG, electrode pop, movement and ocular artefacts) be grouped together in distinguishing them from normal EEG activity. To



Figure 3.1: Real-time EEG Analysis for Event Detection (REACT) prototype device.

the author's knowledge, the work presented here is the first to treat ambulatory EEG for artefacts arising from head movement as a single class in an automatic artefact detection algorithm and as such is a novel approach to the artefact detection problem.

3.2.2 Ambulatory REACT system

Ambulatory REACT (**R**eal-time **E**EG **A**nalysis for Event Dete**CT**ion) is a small form factor system developed by the Efficient Embedded Digital Signal Processing (EEDSP) group at U.C.C. that performs real-time EEG monitoring using a DSP microprocessor. This point-of-care technology is designed for unobtrusive ward and domestic detection of neurological events in adults, whereby the patients wear the device and go about their daily lives as usual (Temko et al., 2010; McEvoy et al., 2010). The system is built around the concept of intelligence-at-the-sensor; whereby the costly requirement of continuously transmitting raw bio-signal data is removed, and only data of relevance (such as seizure alarms) are transmitted. A picture of the REACT prototype is shown in Figure 3.1. REACT is based on a Support Vector Machine (SVM) classifier, and uses information from

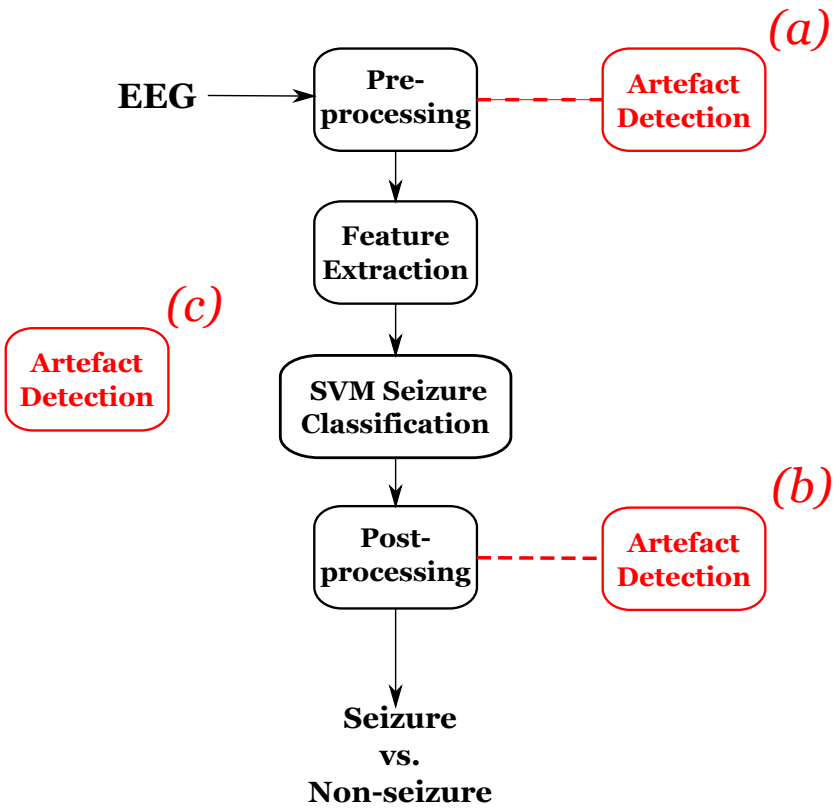


Figure 3.2: Real-time EEG Analysis for Event Detection. This architecture shows three potential methods of incorporating artefact detection into the system. (a) indicates rejection of epochs classified as artefact. (b) indicates classifier fusion of seizure and artefact information. (c) represents artefact annotation, performed separately from the seizure classification task.

a rich set of features to classify EEG epochs as either seizure or non-seizure. The system has been shown to achieve state-of-the-art performance on datasets of adult EEG (Faul et al., 2009; Temko et al., 2010, 2011b); thereby rapidly assisting neurophysiologists in identifying areas of interest in EEG recordings.

The work outlined in this chapter presents the first stages in the development of an artefact detection system to be implemented with ambulatory REACT. Discussions with clinicians at Cork University Hospital as well as reports in the literature have identified head movement artefacts as causing difficulty both in terms of false positive seizure classification and also in obscuring the recorded EEG for subsequent medical examination Barlow (1984, 1986). As outlined in Chapter 2, identifying sections of EEG that are contaminated by artefact would be useful for various applications. Including artefact detection in ambulatory REACT could take a number of guises; Figure 3.2 illustrates how three such methods could be incorporated into REACT. Artefact detection could take the form of detecting epochs that contain head-movement artefact and rejecting these epochs so that the seizure detection classifier is presented with non-artefactual EEG alone. Alternatively, information from the seizure detection and artefact detection classifiers could be combined at the post-processing stage. Finally, artefact detection could be included as an annotation tool for the clinician whereby he/she uses information from the artefact detection classifier when examining the EEG recording.

3.2.3 Artefact detection using supervised machine learning

Artefact detection using supervised machine learning will take the following general framework. Raw EEG signals will be preprocessed so as to remove 50 Hz mains frequency artefact and low-amplitude DC-component of the signal. These steps will be accomplished with finite impulse response (FIR) filters. Each channel of the EEG signal is then segmented into overlapping windows or epochs; a set of relevant features is subsequently extracted from each of these epochs. These features and the labels (i.e. annotations) associated with them are used to train a machine learning classifier to obtain a discriminating boundary between the two classes (artefact and normal EEG) that is optimal in some sense (Abu-Mostafa et al., 2012). That is to say, the machine learning algorithm defines a cost function to represent the separability of the classes and performs an optimisation of the cost function using the training data. The following sections provide details on how this framework was applied to the artefact detection task using a set of 69 EEG features,

and linear discriminant and support vector machines classification models.

3.3 EEG features

Features are quantitative descriptors of the EEG, whose purpose is to show significant changes in value between classes; in this application, to show changes in value during the presence of artefacts arising from head-movement as compared to background EEG activity. These features are chosen to take advantage of differences in the underlying characteristics of background and artefact EEG.

In order to calculate features, the EEG signal is segmented into windows/epochs whose duration is capable of capturing the characteristics of the event to be detected. Head-movement artefacts are comprised of underlying artefacts of differing lengths; in the generalised head-movement detection task, it not clear what window length is most appropriate. Therefore, in this chapter a number of different window lengths are investigated and the corresponding, overall system performance is subsequently evaluated. The segmentation process included overlap of 90% was used so as to increase the amount of data available and has been demonstrated in the literature to be appropriate).

While EEG events can have different shapes, morphologies and timings, there are large similarities and common characteristics between EEG events representing vastly different neurological phenomena (or in the case of artefact, non-neurological events). Consequently, there is significant cross-fertilisation of features between various automated EEG applications and even between different time-series, machine learning applications in general (perhaps most notably, from speech processing to EEG). Feature sets used in EEG applications, ranging from seizure detection to imagined arm movement (as used in BCI) and from sleep cycle classification to Parkinson's detection, tend to be broadly similar or at least contain significant overlap. This is not to say that the events themselves are the same or in some cases even similar, simply that similar features (or groups of features) can be used to differentiate between a broad range of dissimilar events. Taking this into account then, the feature set used in the artefact detection task is drawn primarily from two sources; the REACT and ANSeR neurological event detection systems and several EMG and ocular artefact detection papers in the literature (van de Velde et al., 1999; Gasser et al., 2005).

The full set of 69 features extracted from the EEG is listed in Table 3.1 with detailed

Table 3.1: List of EEG features for head-movement artefact detection.

Time Domain Features
- Line length/ Curve Length (L)
- RMS Amplitude
- Slope
- Activity (1st Hjorth Parameter)
- Mobility (2nd Hjorth Parameter)
- Complexity (3rd Hjorth Parameter)
- Kurtosis
- Skewness
- Nonlinear Energy (N)
- Zero Crossings (Zc)
- Minima and Maxima
- Autoregressive Modelling Error (AR models 1-9)
- Variance of first derivative
- Variance of second derivative
- Zero Crossings of first derivative
- Zero Crossings of second derivative

Frequency Domain Features
- Peak Frequency
- Spectral Edge Frequency(80, 90 and 95)
- Intensity Weighted Mean Frequency (IWMF)
- Intensity Weighted Bandwidth (IWBW)
- Total Power
- Power Bands: 0-2 Hz, 1-3 Hz, 2-4 Hz, 3-5 Hz, 4-6 Hz, 5-7 Hz, 6-8 Hz, 7-9 Hz, 8-10 Hz, 9-11 Hz, 10-12 Hz, 3-15 Hz, 15-30 Hz, 59-61 Hz, 51-64 Hz, 20-30 Hz, 25-64 Hz
- Normalised Power Bands: 0-2 Hz, 1-3 Hz, 2-4 Hz, 3-5 Hz, 4-6 Hz, 5-7 Hz, 6-8 Hz, 7-9 Hz, 8-10 Hz, 9-11 Hz, 10-12 Hz, 3-15 Hz, 15-30 Hz, 59-61 Hz, 51-64 Hz, 20-30 Hz, 25-64 Hz

Entropy-based Features
- SVD Entropy
- Shannon Entropy
- Fisher Entropy
- Spectral Entropy

explanation in Sections 3.3.1, 3.3.2 and 3.3.3. Many features describe the morphology of the EEG in the time domain, with notable differences existing between the two classes. The majority of features are extracted from the frequency domain, where significant differences exist between background EEG and artefact. A further set of features that quantify the underlying structure of the EEG signal using entropy-based measures from information theory are also investigated.

3.3.1 Time domain features

When a clinician or researcher examines an EEG recording, it is time domain features that are observed. These time domain features typically encompass the shape and morphology

of the signal and can include poorly-defined descriptions of the waveform (such as spikiness, uniformity or degree of asymmetry). For automated artefact detection, these qualitative characteristics must be translated into a quantifiable numerical measure. A number of features are extracted from the time domain EEG signal to capture the characteristics that a human EEG reader would use to identify artefact. These features are complimented by additional time domain features (such as time domain models of the EEG signal) that may not be apparent to the human observer, but nevertheless provide useful separation between normal EEG and artefacts. These features are derived from a statistical analysis of the EEG signal within an epoch (or from the first and second derivatives of the EEG signal within an epoch).

Line length/ Curve length

Line length (L) is used as a measure of signal complexity, initially proposed by Esteller et al. (2001) as an indicator of seizure onset. It is similar to the waveform fractal dimension although it has been shown to be more computationally efficient. Line length (sometimes referred to as curve length) is defined for an epoch $\mathbf{x}_j$ as:

$$L(\mathbf{x}_j) = \sum_{i=0}^{n_s} |x_j(i+1) - x_j(i)|. \tag{3.1}$$

Thus, line length is the running sum of distances between consecutive points within the sliding window of size n_s. The discriminating potential of a feature can be visualised by means of a probability density function (pdf), which plots the relative likelihood that the feature will take on a given value. Thus, a feature is discriminative in non-overlapping areas of the pdfs. Figure 3.3 displays probability density functions for the line length of background EEG (blue) and head-movement artefact EEG (green), where it can be seen that the head-movement artefact has a much broader distribution than that of background EEG. This is to be expected as the head-movement artefact is comprised of a number of contributory artefact sources (namely, muscle, ocular, electrode pop and electrode movement artefacts). At lower line lengths the contribution of muscle and ocular artefacts is seen, whereas high amplitude electrode pop explains the broad tail of the distribution at higher line lengths. In this thesis, probability density functions were generated using a random sample of 5000 data points per class.

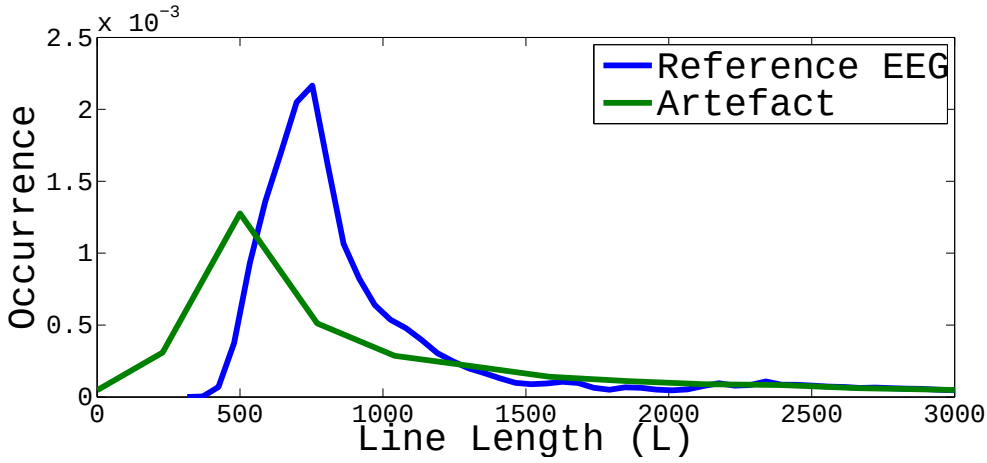


Figure 3.3: Probability density function for line length (L) of background EEG (blue) and head-movement artefact (green).

Root mean square amplitude

The root mean square (RMS) amplitude, or quadratic mean, is a statistical measure of the magnitude of a time varying quantity. The RMS amplitude expresses the mean of the absolute amplitude of an epoch $\mathbf{x}_j$ and is defined as:

$$RMS(\mathbf{x}_j) = \sqrt{\frac{1}{n_s}\sum_{i=1}^{n_s} x_j^2(i)}. \tag{3.2}$$

As artefact EEG is often high-energy, high-amplitude signals, the RMS amplitude aims to capture this trait. This is highlighted in the pdfs in Figure 3.4, where the artefact EEG is distributed farther to the right than that of the background EEG. Once more, as the artefact class comprises a number of different signal types the feature values are more broadly distributed than that of the background EEG.

Slope

The slope of the EEG signal describes its steepness and is calculated by the first derivative $d\mathbf{x}_j(i)$ of the signal, where $d\mathbf{x}_j(i) = \mathbf{x}_j(i) - \mathbf{x}_j(i-1)$. The mean slope of each epoch is given by the cumulative sum over consecutive sample points:

$$\delta_{EEG} = \frac{1}{N-1}\sum_{i=2}^{N} d\mathbf{x}_j(i). \tag{3.3}$$

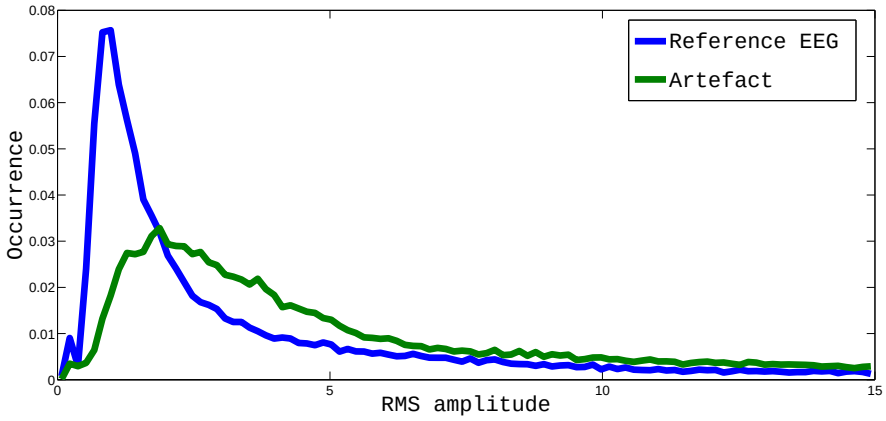


Figure 3.4: Probability density function for RMS amplitude of background EEG (blue) and head-movement artefact (green).

Variance and Hjorth parameters

In probability theory and statistics the variance of a signal σ_j^2 is a measure of how far the numbers in a probability distribution lie from the mean of that distribution. Variance is often referred to as the second central moment and in EEG signal processing is sometimes denoted as activity or the 1st Hjorth Parameter (Hjorth, 1970). Variance or activity is thus given by:

$$Activity(\mathbf{x}_j) = \sigma_j^2 = \frac{1}{n_s}\sum_{i=1}^{n_s}(x_j(i) - \mu_j(\mathbf{x}_j))^2, \tag{3.4}$$

where $\mu_j(\mathbf{x}_j)$ is the sample mean of an epoch $\mathbf{x}_j$ and defined as:

$$\mu_j = \mu(\mathbf{x}_j) = \frac{1}{n_s}\sum_{i=1}^{n_s} x_j(i). \tag{3.5}$$

Owing to the fact that artefacts arise from a number of different source signals, it is anticipated that the variance of artefactual sections of EEG should on average be greater than the variance of normal EEG. This is illustrated in Figure 3.5, where it can be seen that the variance of artefact EEG is more broadly distributed than that of the EEG, with a thick tail depicting more epochs with higher variances. The square root of the variance σ_j^2 is referred to as the standard deviation σ_j. Hjorth (1970) also introduced two further EEG features, mobility and complexity, based on the standard deviation of the first and second derivatives of the EEG signal, respectively. The Hjorth mobility of an epoch is defined as:

$$Mobility(\mathbf{x}_j) = \sigma_{\Delta j}/\sigma_j, \tag{3.6}$$

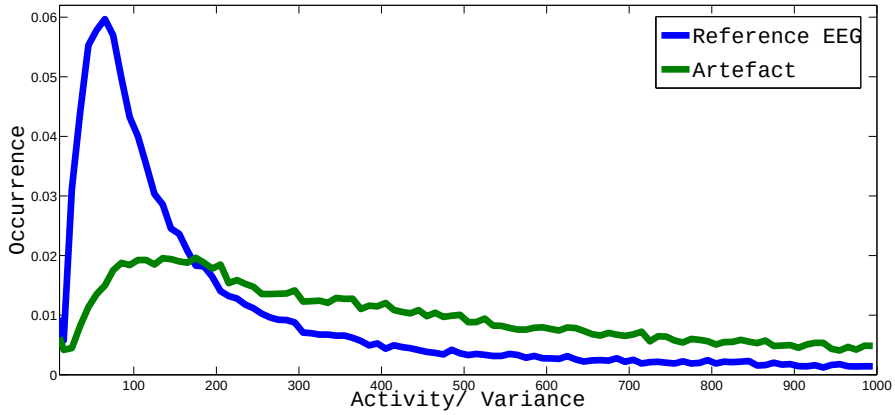


Figure 3.5: Probability density function for activity or variance of background EEG (blue) and head-movement artefact (green).

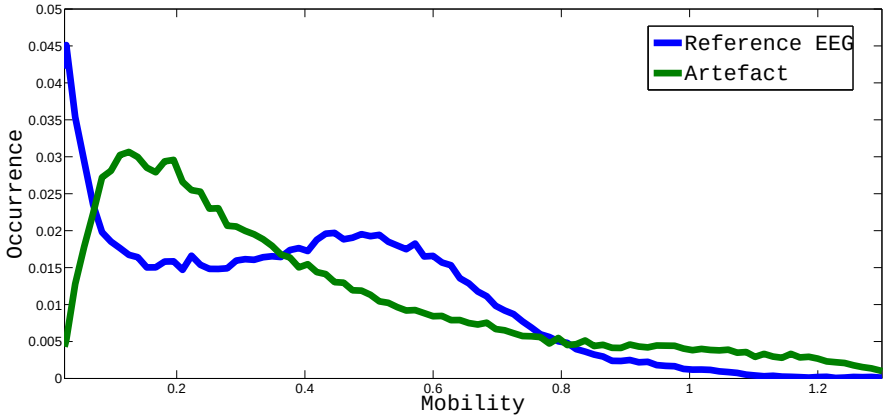


Figure 3.6: Probability density function for the mobility of background EEG (blue) and head-movement artefact (green).

where $\sigma_{\Delta j}$ is the standard deviation of the first derivative of an epoch. The pdfs for the mobility of each class are displayed in Figure 3.6. As with the other features, the artefact EEG distribution has a long tail, with a single broad peak.

The complexity of an epoch is defined as:

$$Complexity(\mathbf{x}_j) = \frac{\sigma_{\Delta 2j}/\sigma_{\Delta j}}{\sigma_{\Delta j}/\sigma_j}, \tag{3.7}$$

where $\sigma_{\Delta 2j}$ is the standard deviation of the second derivative of the epoch.

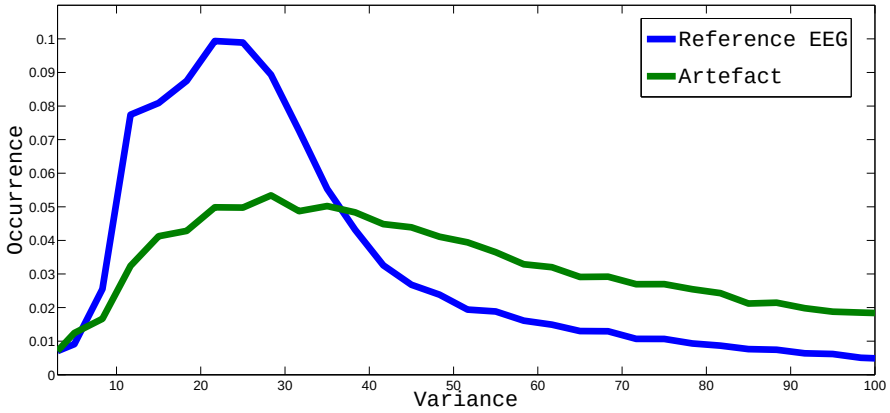


Figure 3.7: Probability density function for the variance of the 1^{st} derivative of background EEG (blue) and head-movement artefact (green).

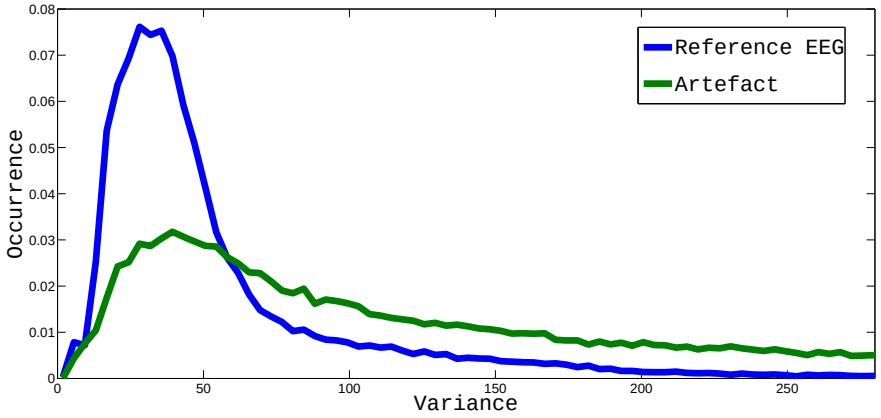


Figure 3.8: Probability density function for the variance of the 2^{nd} derivative of background EEG (blue) and head-movement artefact (green).

Variance of first and second derivatives

The variance of the first ($\sigma_{\Delta j}$) and second ($\sigma_{\Delta 2j}$) derivatives were used by Thomas et al. (2010) for neonatal seizure detection. They are included here as features in the artefact classification task. Probability density functions for the variance of the 1^{st} and 2^{nd} derivatives are displayed in Figures 3.7 and 3.8. As for the variance, the artefact EEG is more broadly distributed than that of the background EEG, with long tails at higher values.

Skewness

In probability theory and statistics, skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable. A negative skew indicates that the tail on the left side of the probability density function (pdf) is longer than that of the right side, with the bulk of the values lying to the right of the mean. Conversely, a positive skew indicates that the tail on the right side is longer than that of the left and the bulk of the values lie to the left of the mean. Zero skewness indicates that the values are relatively evenly distributed either side of the mean, typically (but not necessarily) implying a symmetric distribution. Skewness is often referred to as the third central moment and is defined as:

$$Skewness(\mathbf{x}_j) = \frac{\frac{1}{n_s}\sum_{i=1}^{n_s}(x_j(i) - \mu_j)^3}{(\frac{1}{n_s}\sum_{i=1}^{n_s}(x_j(i) - \mu_j)^2)^{3/2}}. \tag{3.8}$$

A probability density function is not displayed here for skewness as the feature did not exhibit significant discrimination on its own.

Kurtosis

Kurtosis, often referred to as the fourth central moment, is a measure of the "peakedness" of a probability density function and is defined as follows:

$$Kurtosis(\mathbf{x}_j) = \frac{\frac{1}{n_s}\sum_{i=1}^{n_s}(x_j(i) - \mu_j)^4}{(\frac{1}{n_s}\sum_{i=1}^{n_s}(x_j(i) - \mu_j)^2)^2}. \tag{3.9}$$

A high kurtosis distribution has a sharp peak and long, fat tails, while a low kurtosis distribution has a more rounded peak and short, thin tails. Unlike many of the other displayed features, the pdfs of normal and artefact EEG overlap significantly for this feature, as seen in Figure 3.9. However, as will be discussed in Section 3.4 if the marginal class distributions are overlapping this does not necessarily imply that the feature cannot provide useful discrimination when combined with additional features in a higher dimensional space.

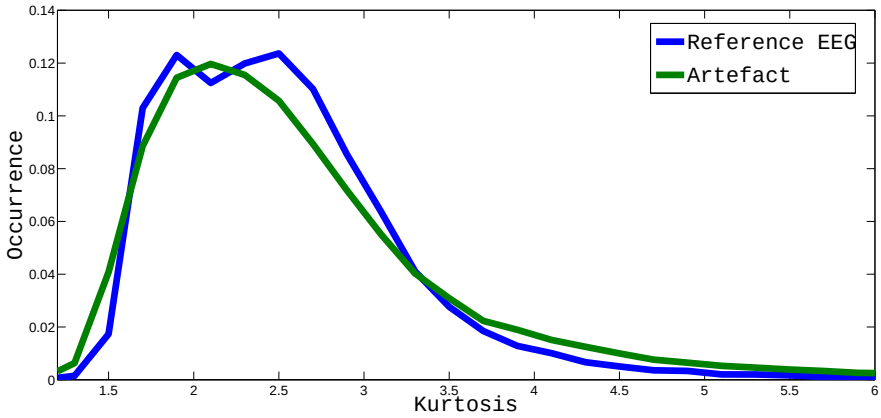


Figure 3.9: Probability density function for the kurtosis of background EEG (blue) and head-movement artefact (green).

Nonlinear energy (NLE)

Non-linear energy (NLE) is a function of the amplitude of a signal, and the change of that amplitude and is defined as:

$$NLE(\mathbf{x}_j) = \frac{1}{n_s - 2} \sum_{i=2}^{n_s-1} x_j(i)^2 - x_j(i-1)x_j(i+1). \tag{3.10}$$

NLE was introduced by D'Alessandro et al. (2003) as a feature in an epileptic seizure prediction algorithm. Greene et al. (2008), Thomas et al. (2010) and Temko et al. (2011b) have found NLE to be useful in neonatal seizure detection also. The pdfs of NLE are displayed in Figure 3.10 where it can be seen that the artefact EEG is more uniformly distributed than the background EEG, where low NLE values are most common.

Number of zero crossings (Zc) and its derivatives

The number of zero crossings (Zc) is the number of times within an epoch that the EEG signal crosses the x-axis. The number of zero crossings of the 1st derivative of the EEG corresponds to the number of local maxima and minima of the EEG. The number of zero crossings of 2nd derivative corresponds to the number of times that the 2nd derivative of the EEG signal crosses the x-axis within an epoch. In Figures 3.11 and 3.12, it can be seen that the pdfs of normal and artefact EEG for the number of zero crossings of the 1^{st} and 2^{nd} derivatives overlap considerably. However, as will be seen in Sections 3.4 and 3.5,

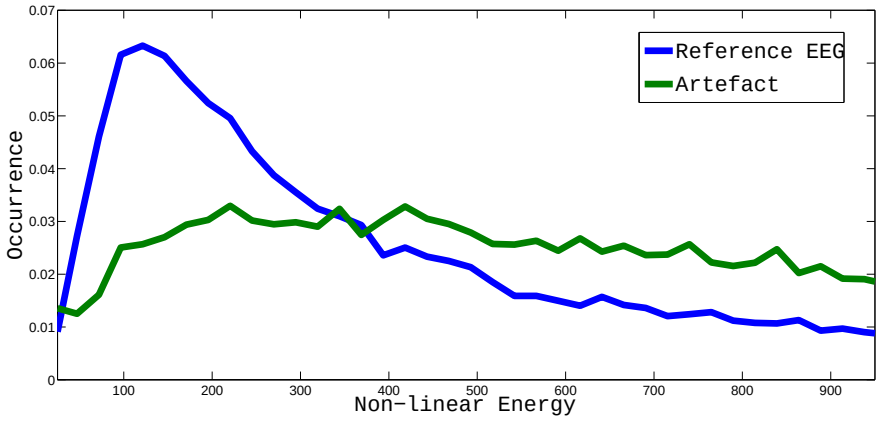


Figure 3.10: Probability density function for the non-linear energy of background EEG (blue) and head-movement artefact (green).

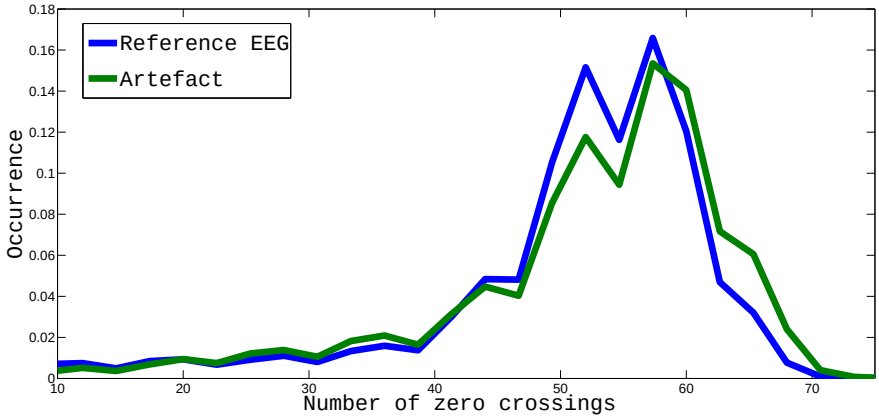


Figure 3.11: Probability density function for the number of zero crossings of the 1^{st} derivative of background EEG (blue) and head-movement artefact (green).

when combined with other features, these features do improve classification performance in the artefact detection task.

Number of inactive samples

The number of inactive samples within an epoch $\mathbf{x}_j$ is defined as the number of samples for which there is very little change in the EEG amplitude. This was calculated by applying a threshold of 0.01 to the absolute value of the derivative of the EEG signal.

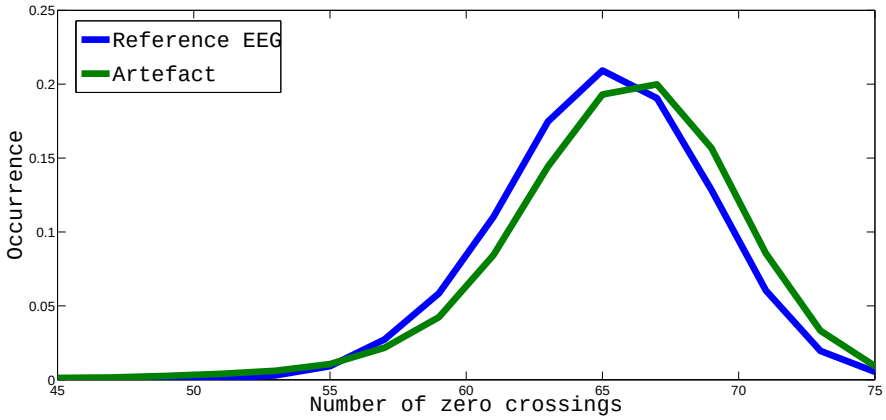


Figure 3.12: Probability density function for the number of zero crossings of the 2^{nd} derivative of background EEG (blue) and head-movement artefact (green).

Autoregressive modelling error

An autoregressive (AR) model can be used for prediction in a correlated time series. A variable $\hat{x}(k)$ in a correlated time series can be predicted from previous observations in the series by:

$$\hat{x}_j(k) = \sum_{i=1}^{n_p} \phi_j(i)\hat{x}_j(k-i) + \epsilon_k, \tag{3.11}$$

where $\phi_j(i)$ are the n_p parameters of the AR model and ϵ_k is a zero mean, white noise term accounting for the error in each prediction step. The $\phi_j(i)$ parameters of the AR model are estimated over the first half of the epoch $\mathbf{x}_j$. The AR model is fit to the data over the first half of the epoch using the Yule-Walker method (Kay, 1988) and the model is used to perform one step ahead prediction on the second half of the epoch. The percentage error is then given by:

$$AR_{fit}(\mathbf{x}_j) = 100\left(1 - \frac{\sum_{k=\frac{n_s}{2}+1}^{n_s} |x_j(k) - \hat{x}_j(k)|}{\sum_{k=\frac{n_s}{2}+1}^{n_s} |x_j(k) - \overline{\mathbf{x}}_j|}\right), \tag{3.12}$$

where:

$$\overline{\mathbf{x_j}} = \frac{1}{n_s/2} \sum_{k=\frac{n_s}{2}+1}^{n_s} x_j(k). \tag{3.13}$$

A total of 9 features are generated using this approach, corresponding to models of orders 1 to 9. Figure 3.13 illustrates pdfs for the autoregressive model fit for 1^{st} order AR model.

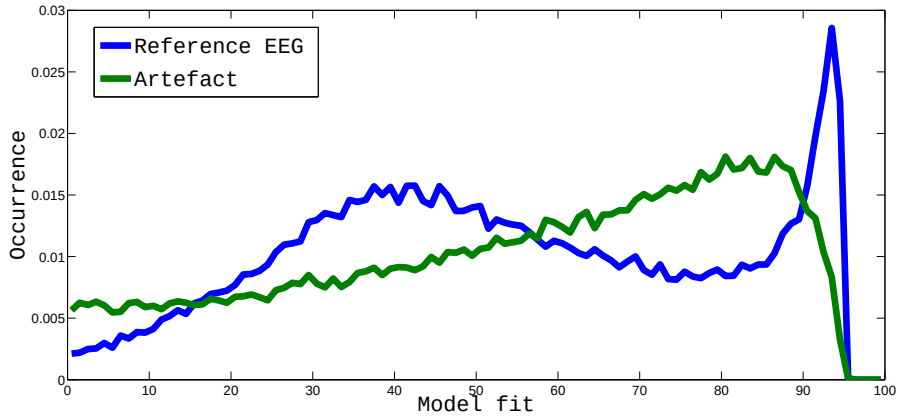


Figure 3.13: Probability density function for the autoregressive model fit (1^{st} order) of background EEG (blue) and head-movement artefact (green).

3.3.2 Frequency domain features

When a clinician or researcher examines an EEG trace, they explicitly (via a fast Fourier transform (FFT) computed in the visualisation software) or implicitly (by observing the periodicity of events in the EEG signal), utilise information from the frequency domain. Consequently, features from the EEG's frequency domain are used to quantify changes in the spectrum of the EEG during the presence of artefact. The power spectral density (PSD) of an epoch is obtained using a 128 point FFT. The FFT gives an output of n_s complex coefficients, which are converted to real values by taking the absolute value of the coefficients. The spectrum of an EEG epoch $\mathbf{x}_j$ can be expressed in vector form as frequency coefficients $\mathbf{a}_j = [a_j(0), a_j(1), ..., a_j(i), ..., a_j(\frac{n_s}{2})]$ where $a_j(i)$ is the amplitude of a sinusoid of frequency $i\frac{f_s}{n_s}$.

The following frequency domain features are then derived from the PSD of each epoch:

Peak frequency

Peak frequency is defined as the frequency corresponding to the largest amplitude in the power spectral density (PSD). It is the dominant frequency component in the EEG signal for that epoch and should characterise to some degree the underlying source signal.

$$f_{peak}(\mathbf{x}_j) = i_{peak}\frac{n_s}{f_s}, \quad \text{where} \quad i_{peak} = \text{argmax } a_j(i). \tag{3.14}$$

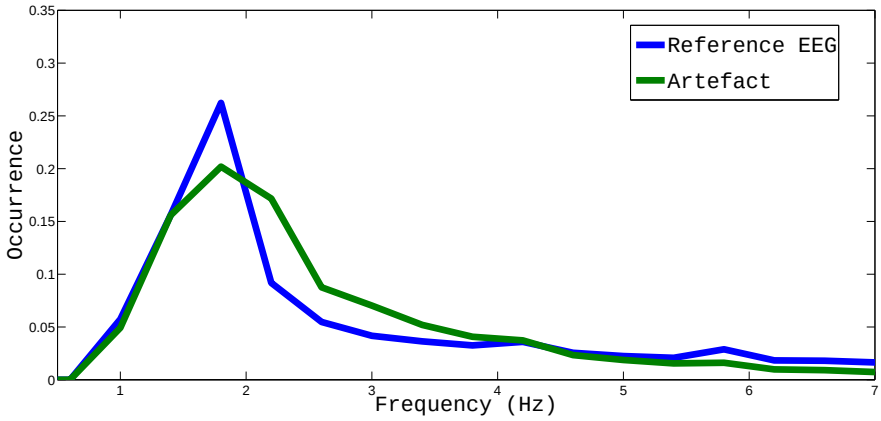


Figure 3.14: Probability density function for the SEF_{80} of background EEG (blue) and head-movement artefact (green).

Spectral edge frequency(80, 90 and 95)

Spectral edge frequency (SEF) is defined as the frequency under which a certain percentage of the power in the PSD lies. In this work, three SEF features are used, corresponding to 80%, 90% and 95% of the power in the PSD. Accordingly, higher SEFs should be influenced by head-movement artefact arising from EMG sources. This is illustrated in Figure 3.14 where the pdf of the artefact EEG is located slightly toward higher frequency values than that of the normal EEG.

Intensity weighted mean frequency (IWMF)

The intensity weighted mean frequency f_m is the average frequency from the frequency spectrum, and defined as:

$$f_m = \frac{\sum_{i=0}^{\frac{N}{2}-1} p_i \; i \; df}{\sum_{i=0}^{\frac{N}{2}-1} p_i}, \tag{3.15}$$

where i is the frequency bin number, p_i is the estimated spectral power in the bin and $df = f_s/N$, with f_s being the sampling frequency and N the total number of frequency bins. The IWMF corresponds to the expected frequency value in an EEG epoch $\mathbf{x}_j$.

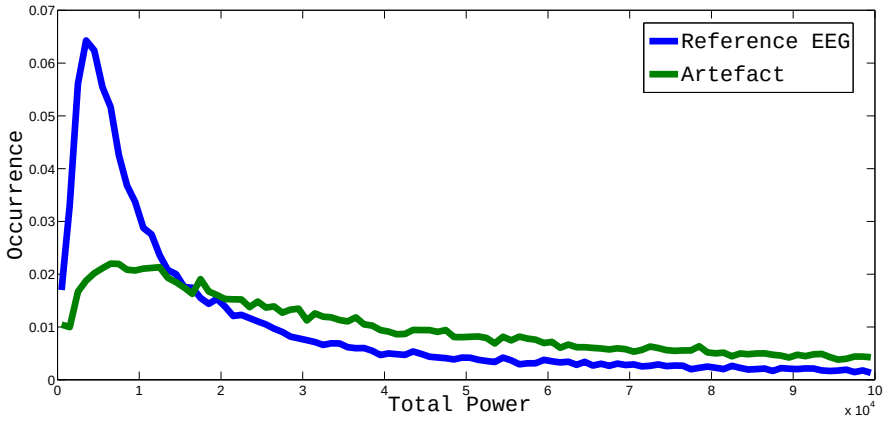


Figure 3.15: Probability density function for the total power in the 0 to 12 Hz range of background EEG (blue) and head-movement artefact (green).

Intensity weighted bandwidth (IWBW)

The intensity weighted bandwidth b_w is defined as:

$$b_w = \sqrt{\frac{\sum_{i=0}^{\frac{N}{2}-1} p_i (f_m - i\ df)^2}{\sum_{i=0}^{\frac{N}{2}-1} p_i}}. \tag{3.16}$$

Total power

The total power refers to the sum of power in all bins of the PSD between 0 and 12 Hz:

$$P_{total}(\mathbf{x_j}) = \sum_{i=0}^{12n_s/fs} p_j(i), \tag{3.17}$$

where $p_j(i)$ is the power in bin i of epoch $\mathbf{x}_j$. This feature was taken from epilepsy (D'Alessandro et al., 2003; Saab and Gotman, 2005; Faul et al., 2009) and neonatal seizure detection (Greene et al., 2008; Temko et al., 2009; Thomas et al., 2010) papers and refers to the total power in the frequency of typically normal EEG. As shown in Figure 3.15 the power in artefact EEG tends be distributed more uniformly than that of the background EEG which has a peak at 0.75 Hz.

Power bands and normalised power bands

Powers in specific EEG sub-bands are widely used features for EEG analysis in epilepsy research (Shoeb et al., 2004), BCI (Pfurtscheller et al., 1997) and neonatal seizure detection (Thomas, 2011). In this study, power in frequency bands of 2 Hz width are extracted from the PSD calculation; i.e. power in 0-2 Hz, 1-3 Hz, 2-4 Hz, 3-5 Hz, 4-6 Hz, 5-7 Hz, 6-8 Hz, 7-9 Hz, 8-10 Hz, 9-11 Hz, 10-12 Hz. The power in additional frequency bands are also utilised as features, owing to their usefulness in other papers on muscle artefact detection; these bands are 3-15 Hz, 15-30 Hz, 59-61 Hz, 51-64 Hz, 20-30 Hz, 25-64 Hz.

In addition to the power in each sub-band, the normalised powers in these sub-bands were also used as features. The normalised power in a sub-band reflects the proportion of overall signal power existing in a given sub-band, and are calculated by dividing the power in a sub-band by the total power in the signal for that epoch.

3.3.3 Entropy-based features

In information theory, entropy is a measure of the uncertainty in a random variable. Shannon (1949) introduced the concept of entropy in the context of digital communication but it has since proved an effective tool in the prediction and characterization of other signals. Consequently, entropy as introduced by Shannon (1949) as well as other information measures are utilised here as features to characterise EEG of differing types.

Shannon entropy

Shannon entropy is a measure in information theory for estimating the uncertainty of an outcome (Shannon, 1949). It is the average unpredictability in a random variable, which is equivalent to its information content. To calculate Shannon entropy, the signal must first be represented as a discrete distribution. This is performed here by approximating the probability mass function by a 16-bin histogram. The Shannon entropy of the j^{th} epoch is thus defined as:

$$H_{Sh}(\mathbf{x}_j) = -\sum_{i=1}^{16} p_i(\mathbf{x}_j)\ log\ p_i(\mathbf{x}_j), \tag{3.18}$$

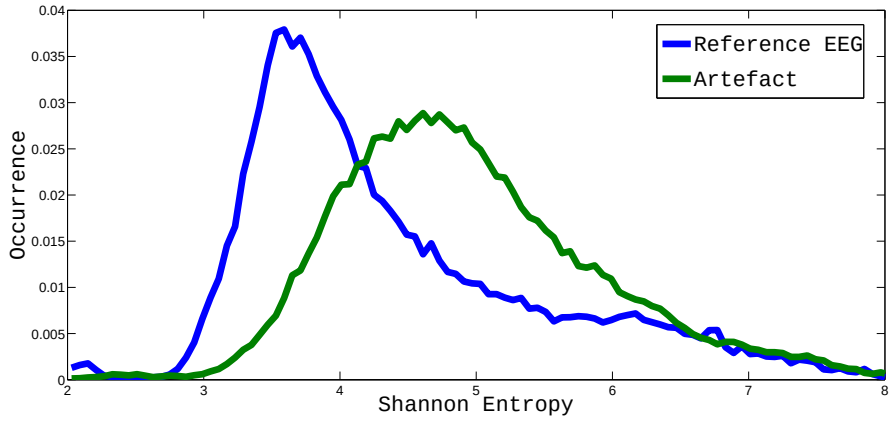


Figure 3.16: Probability density function for the Shannon entropy of background EEG (blue) and head-movement artefact (green).

where $p_i(\mathbf{x}_j)$ is the magnitude of each bin. If the entropy of $\mathbf{x}_j$ is zero, the observer is certain of the future value of $\mathbf{x}_j$. Higher values of entropy then indicate increased uncertainty. In Figure 3.16 it can be seen that the artefact EEG exhibits higher entropies (i.e. more disorder) than that of the normal background EEG.

Spectral entropy

Where the Shannon entropy is used to quantify the order in the EEG signal, spectral entropy is a measure of the order in the frequency spectrum of the EEG:

$$H_{Spec}(\mathbf{x}_j) = -\frac{1}{log\ N}\sum_{i=1}^{N} z_i\ log\ z_i, \tag{3.19}$$

where i is a frequency index and z_i is a normalised power spectral density $S(\omega_i)$:

$$z_i = \frac{S(\omega_i)}{\sum_{j=1}^{N} S(\omega_j)}. \tag{3.20}$$

SVD entropy

Singular Value Decomposition (SVD) is a measure of the complexity of a signal, often used to obtain information about quasi-periodic signals in noise. The SVD algorithm decomposes a matrix such that:

$$\mathbf{A} = \mathbf{USV}^T \tag{3.21}$$

where $\mathbf{A}$ is the input matrix, where $\mathbf{U}$ and $\mathbf{V}$ have orthogonal columns such that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, with $\mathbf{I}$ being the identity matrix and $\mathbf{S}$ is a diagonal vector of singular values. The singular values in $\mathbf{S}$ refer to the most significant underlying components in the signal. The number of singular values varies with the complexity of the signal, with an increase in signal complexity leading to a larger number of singular values. The number of significant singular values $\zeta_1 ... \zeta_{d_E}$ can be obtained using Rissanen's Minimum Description Length algorithm (Roberts et al., 1999).

The SVD entropy calculates the entropy in the singular spectrum (Roberts et al., 1999). By performing SVD for an epoch as described in Equation 3.21, the singular values $\zeta_1 ... \zeta_{d_E}$ can be found. The SVD entropy is thus:

$$H_{SVD} = -\sum_{i=1}^{d_E} \hat{\zeta}_i \; log_2 \; \hat{\zeta}_i, \tag{3.22}$$

where d_E is the singular dimension given by Rissanen's Minimum description length, and where $\hat{\zeta}_i$ is the normalised singular values such that

$$\hat{\zeta}_i = \frac{\zeta_i}{\sum_j \zeta_j}. \tag{3.23}$$

SVD entropies should be lower for quasi-periodic signals such as EEG baseline oscillations due to movement.

Fisher entropy

The Fisher information is calculated from the singular values of the EEG to describe the shape of the singular spectrum.

$$I_{Fisher}(\mathbf{x}_j) = \sum_{i=1}^{d_E - 1} \frac{(\hat{\zeta}_{i+1} - \hat{\zeta}_i)^2}{\hat{\zeta}_i}. \tag{3.24}$$

3.4 Feature reduction and linear discriminant classification

The features outlined in Section 3.3 describe varying levels of separation between background and artefact EEG classes. A statistical pattern recognition classifier then attempts to use these features to find some optimal separating hyperplane such that features ex-

tracted from normal EEG lie on one side of the hyperplane and features extracted from artefact EEG lie on the other. Linear discriminant classifiers have a very low computational requirement and generally provide reasonable results, making them suitable for proof-of-concept experiments. Additionally, the low computational burden makes LDCs ideal for online classification; in particular LDCs have been widely used in BCI applications (Allison et al., 2007; Lotte et al., 2007). Therefore, the premise that a generalised head-movement artefact class can be used to detect artefacts introduced into the EEG by head-movements, is tested with linear discriminant classification.

The majority of papers in the literature extract features from one second windows of EEG. However, this has rarely been supported with compelling evidence as to why this window length was chosen. As head-movement artefacts occur with differing lengths, it is not clear what window length best captures the signal characteristics. In order to investigate what window length would be most appropriate for classifying head-movement artefact, this classification task was performed for window lengths of 0.25, 0.5, 0.75, 1, 1.5, 2, 3, and 4 seconds.

3.4.1 Feature reduction using mutual information evaluation function

For most machine learning classifiers, a large feature set may be detrimental to the performance of the classifier, in particular when the number of data points is limited. This "curse of dimensionality", which implies that complexity of the classification task increases exponentially with the dimension of the data, means that in practice it may be beneficial to limit the number of features to only those which are most discriminative for the classification task (Abu-Mostafa et al., 2012). There exists a wide range of methods available to reduce the size of the feature set, broadly categorised as filter techniques, wrapper methods and embedded techniques (Saeys et al., 2007). Generally speaking, filter techniques are independent of the classifier, whereas wrapper and embedded methods require a classifier to be chosen before feature reduction is performed. Filter techniques assess the relevance of features by examining intrinsic properties of the data, usually by calculating some feature relevance score (such as correlation or mutual information) that characterises the information between the feature and the class labels. Low-scoring features are then removed, and the remaining features are presented to the chosen classification algorithm. Filter techniques are those independent of the chosen classifier, and offer a general picture as to what features might be useful for the classification task, irrespective of classification

algorithm. As the feature selection was performed at the proof-of-concept stage where a final classification algorithm is yet to be chosen, a filter technique was employed.

A multi-variate filter technique, mutual information evaluation function (MIEF), was chosen (Al-Ani and Deriche, 2002). Mutual Information is used here as a measure of usefulness of individual features in distinguishing between normal EEG and head movement artefact. MIEF chooses the feature with highest mutual information between the feature and the class labels. Additional features were subsequently ranked by information gain based on how much additional information they provided to the classification problem. MIEF is therefore more useful than simply using mutual information alone to rank features as it takes redundancy between features into account, obtaining the best group performance of features. Additionally, it is better than using simple correlation to rank features, as the correlation takes only linear connections between series into account, and classification algorithms often exploit non-linear similarities in the data. The idea of selecting EEG features based on Mutual Information (MI) using the Mutual Information Evaluation Function (MIEF) proposed by Deriche and Al-Ani (2001) has previously been demonstrated for seizure detection systems by Faul (2007). The MI $I(X;Y)$ between random variables X and Y measures the amount of information in X that can be predicted when Y is known. If X and Y are continuous, then:

$$I(X;Y) = H(X) - H(X|Y) = \int P_{XY}(x,y)\ log\frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}\ dxdy, \tag{3.25}$$

where $H(X)$ is the entropy of X and $H(X|Y)$ is the conditional entropy of X given Y. Converting to discrete data, by dividing the XY plane into boxes of size $\Delta_x\Delta_y$, gives MI as:

$$I(X;Y) = \sum_{r_x}\sum_{r_y} P_{XY}(r_x,r_y)\ log\frac{P_{XY}(r_x,r_y)}{P_X(r_x)P_Y(r_y)}, \tag{3.26}$$

where r_x and r_y are the discrete levels of X and Y respectively.

The MIEF algorithm proposed by Al-Ani and Deriche (2002) selects features based on their MI with the output $I(C;f_i)$, their MI with the previously selected features $I(f_i;f_j)$ and the joint MI of the test feature and the previously selected features with the output $I(C;\{f_i,f_j\})$. The MIEF algorithm can thus be described as follows:

1. For each feature $f_i \in \mathcal{L}$, choose the feature that has the maximal $I(C;f_i)$.
   (a) Set $\mathcal{K} \leftarrow \{f_i\}$.

(b) Set $g(\mathcal{K}) \leftarrow I(C; f_i)$.

2. For each feature $f_i \in \mathcal{L}, f_i \notin \mathcal{K}$, compute:

$$m(f_i) = g(\mathcal{K}) + \lambda I(C; f_i), \tag{3.27}$$

and choose the feature that maximises $m(f_i)$.

(a) Set $\mathcal{K} \leftarrow \mathcal{K} \cup \{f_i\}$.

(b) Set $g(\mathcal{K}) = m(f_i)$.

3. If $|\mathcal{K}| < |\mathcal{L}|$, go to step 2.

4. $g(\mathcal{L}) = g(\mathcal{K})$

3.4.2 MIEF results

A feature set of the 12 best-performing features was chosen using MIEF for each window length. A subset of 12 features is chosen as it corresponds to the levelling-off of information gain for many of the window lengths investigated. That is to say, adding more features did not substantially increase the mutual information between the feature set and the class labels. 12 of the best performing features, for window length of one second are listed in Table 3.3a, where $I(c, f)$ is the Mutual information between the feature and the class labels, and the change in $m(f)$ is a measure of the information gained by adding an additional feature. Similarly, the 12 best performing features for window length = 1.5 seconds are displayed in Table 3.3b.

The mutual information calculations performed as part of the MIEF algorithm illustrate that the feature set appears to be a good fit with the target classes. Additionally, several features which did not show high mutual information between it and the target class, did contain information that other features did not hold, and thus added to the classifier performance. This is particularly prescient in dealing with a generalised movement artefact class; as features which exhibit good correlation with component signals may perform poorly for the head movement class as a whole. The feature may therefore contribute significant information in terms of describing one of the component artefact signals, while testing poorly for describing the head movement. This is illustrated in the data by the power in the frequency bands 59- 61 Hz and 51- 64 Hz, which had poor correlation between feature and the target class (0.0024196, and 0.001474 for window length = 0.75) but ranked

high (18 and 20) in the MIEF algorithm as there was some information gain associated with their inclusion.

There was no obviously bad feature, that consistently ranked low, or indeed ranked low in terms of both mutual information and information gain.

There was variation in the feature set chosen for each window length. This is understandable given the fact that the features may each capture signal characteristics that are most evident over different signal time lengths. Additionally, in many cases there may have been very little difference in the information gain value assigned to several features at a given time. If one feature was chosen over another, the second feature would likely be relegated down the table of chosen features in the next pass of the MIEF algorithm. Thus, if MIEF is performed for a different window length, where the information gain results at a given split is slightly altered, the MIEF results could alter significantly. However, despite that caveat, the following features performed well for most window lengths: autoregressive fit (several models), SVD entropy, Shannon entropy, zero crossings and line length. This is highlighted in Tables 3.3a and 3.3b. From this, it can be deduced that this feature set is robust to differences in window length, and that these features tend to provided different sources of information for the classification task. The results from the MIEF algorithm show that while there was a difference between the features chosen, the mutual information between individual features and the target was high (>0.05) for many features. This can be explained in part by the fact that while a lot of the features perform well in terms of the mutual information between it and the target class, many did not necessarily add significant information gain. Thus if one feature is ranked highly, similar features tend to be ranked toward the bottom of the MIEF table. Note however, that this is not the case for the autoregressive model features; AR models of different orders can produce significantly different results.

3.4.3 Linear discriminant classification

A linear discriminant (LD) is a function that takes an input vector $\mathbf{x}_j$ and assigns it to one of k classes, $w_c = w_1, ..., w_k$; the function or decision surface, $g(\mathbf{x})$ is a hyperplane. Linear discriminant classification then is a statistical pattern recognition method that finds a linear combination of features that characterises or separates two or more classes of events. As discussed in Section 2.5, the decision surface is found by training the linear discriminant classifier (LDC) on labelled (i.e. annotated) data. In this work, only 2-

Table 3.2: Top performing EEG features for the artefact detection task as ranked by MIEF for window lengths of (a) 1 second and (b) 1.5 seconds

Feature	I(c,f)	m(f)
Line length	0.17015	-
Shannon Entropy	0.13372	0.26378
SVD Entropy	0.067516	0.3313
Autoregressive fit 1	0.090545	0.42184
Autoregressive fit 2	0.070618	0.49246
Activity	0.0664411	0.55887
Autoregressive fit 4	0.06312	0.62199
Autoregressive fit 5	0.06239	0.68438
Zero Crossings	0.062222	0.7466
Autoregressive fit 7	0.062065	0.80867
Autoregressive fit 9	0.06124	0.86991
Autoregressive fit 3	0.059916	0.92983

(a) MIEF for window length 1 second

Feature	I(c,f)	m(f)
SVD Entropy	0.2489	-
Shannon Entropy	0.0534	0.30192
Zero crossings of first derivative	0.11608	0.41797
Autoregressive fit 1	0.1372	0.5079
Autoregressive fit 2	0.14133	0.57217
Autoregressive fit 3	0.13823	0.62502
Spectral Edge Frequency 95	0.045549	0.67056
Autoregressive fit 5	0.14131	0.77475
Autoregressive fit 8	0.13217	0.84829
Zero Crossings	0.041488	0.88978
Line length	0.036975	0.92675
Autoregressive fit 4	0.14514	0.96465

(b) MIEF for window length 1.5 seconds

class classification is considered, with classes corresponding to head-movement artefact and background EEG. The following paragraphs provide some theoretical background of linear discriminant classification derived in a Bayesian framework, i.e. using a generative approach.

LDC theoretical background

Bayesian decision theory is a fundamental statistical approach to the linear discriminant method of pattern classification. This approach is based on quantifying the trade-offs between classification decisions using probability and the costs that accompany such decisions. Bayesian decision theory makes the assumption that the decision problem is posed in probabilistic terms, and that all of the relevant probability values are known (Duda et al., 1995). As such, deciding which class c that a test point $\mathbf{x}_j$ belongs to is based on which class has the highest posterior probability $P(w_c|\mathbf{x}_j)$, that is:

$$g(\mathbf{x}_j) = P(w_c|\mathbf{x}_j), c = 1, 2. \tag{3.28}$$

Bayes' theorem states that the posterior probability of a feature vector $\mathbf{x}_j$ belonging to class w_c is related to the prior probabilities of both classes and the conditional probability of the vector $\mathbf{x}_j$ given the classes:

$$P(w_c|\mathbf{x}_j) = \frac{p(\mathbf{x}_j|w_i)P(w_c)}{\sum_{c=1}^{2} p(\mathbf{x}_j|w_c)P(w_c)}. \tag{3.29}$$

Since $\sum_{c=1}^{2} p(\mathbf{x}_j|w_c)P(w_c)$ is the same for both classes, the denominator can be omitted. As the decision boundary is not unique, the decision is unchanged by multiplying by a constant or shifted by a constant. That is, if $g(\mathbf{x})$ is replaced by $f(g(\mathbf{x}))$, where $f(.)$ is a monotonically increasing function, the result is also unchanged. Thus, the discriminant function is:

$$g(\mathbf{x}) = ln\ p(\mathbf{x}_j|w_c) + ln\ P(w_c). \tag{3.30}$$

Since the likelihood $p(\mathbf{x}_j|w_c)$ is assumed Gaussian it can be represented by:

$$p(\mathbf{x}_j|w_c) \sim \mathcal{N}(\mu_c|\mathbf{\Sigma}_c) = \frac{1}{(2\pi)^{d/2}|\sum_c|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_c)^T\Sigma_c^{-1}(\mathbf{x}-\mu_c)}, \tag{3.31}$$

where μ_c and $\mathbf{\Sigma}_c$ are respectively, the two-dimensional mean vector and 2-by-2 dimension between-class covariance matrix, given by:

$$\mathbf{\Sigma}_c = \sum (\mathbf{x}_j - \mu_c)(\mathbf{x}_j - \mu_c)^T, \tag{3.32}$$

and $|\mathbf{\Sigma_c}|$ is the determinant of the covariance matrix. Thus:

$$g(\mathbf{x}_j) = ln\ P(w_c) + ln\left(\frac{1}{(2\pi)^{d/2}|\sum_c|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_c)^T\mathbf{\Sigma}_c^{-1}(\mathbf{x}-\mu_c)}\right) \tag{3.33}$$

$$= ln\ P(w_c) - \frac{1}{2}ln(2\pi) - \frac{1}{2}ln(|\mathbf{\Sigma}_c|) - \frac{1}{2}(\mathbf{x}-\mu_c)^T\mathbf{\Sigma}_c^{-1}(\mathbf{x}-\mu_c). \tag{3.34}$$

As linear discriminant classifiers assume that the covariances are equal ($\mathbf{\Sigma}_c = \mathbf{\Sigma}$), and discarding all terms that are not dependent on w_c, then:

$$g(\mathbf{x}_j) = ln\ P(w_c) - \frac{1}{2}\mu_c^T\mathbf{\Sigma}^{-1}\mu_c + \mu_c^T\mathbf{\Sigma}^{-1}\mathbf{x}_j \tag{3.35}$$

$$= \mathbf{w}^T\mathbf{x_j} + w_0. \tag{3.36}$$

The discriminant function $g(\mathbf{x})$ is thus a linear combination of the components of $\mathbf{x}_j$, where $\mathbf{x}_j$ is the weight vector and w_0 is the bias or threshold weight.

A two-category linear classifier (e.g. artefact vs. normal EEG) implements the following decision rule:

$$g(\mathbf{x}) \begin{cases} > 0 & \text{Decide } w_1 \\ < 0 & \text{Decide } w_2 \end{cases} \tag{3.37}$$

The equation $g(\mathbf{x}) = 0$ defines the decision surface that separates points assigned to

w_1 from those assigned to w_2. This separating hyperplane is then obtained by seeking the projection that maximises the distance between the means of the two classes, while minimising the inter-class variance.

As there are no parameters to change in the LD classifier, the performance of the LDC is measured by the sensitivity and specificity values (Section 2.9.4). An additional metric, *accuracy*, is also used here, to incorporate both sensitivity and specificity in one value; the accuracy is defined as the mean of the sensitivity and specificity values.

LDC model selection

In order to test the premise that a generalised head-movement artefact class could be used to detect artefacts introduced into the EEG by head-movements, a linear discriminant classification algorithm was used. LD classifiers were trained and tested on the head-movement artefact data set described in Section 2.9.1 using 5-fold cross-validation (Kohavi, 1995). Thus, LDCs were trained using four fifths of the data from all participants and tested on the remaining fifth of data points. This process was then repeated five times, ensuring that the classification results were not overly optimistic or pessimistic due to a biased train/test split. This 5-fold cross validation was performed for each of the following window lengths: 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4 seconds. For each window length, the 12 highest-ranking features from the MIEF algorithm were chosen. This training and testing process was repeated 20 times per window length (to account for potential biases in the randomly selected data), and the resulting mean sensitivity and specificity values were reported.

3.4.4 LDC performance

Specificity and sensitivity results from the LDC are shown in Figure 3.17, indicating that the feature set is useful in separating between normal EEG and artefacts induced by head movement. Accuracy is the mean value of sensitivity and specificity and is included here as a single-figure indication of classification performance. Classification performance was best for those window lengths in the middle of the examined range, i.e., for window lengths of 0.5, 0.75, 1, and 1.5 seconds, where classifier accuracies were highest. For window lengths greater and shorter than this, a drop in classification performance was seen. In particular, sensitivity was highest for window lengths of 1.5 seconds, and specificity was highest for

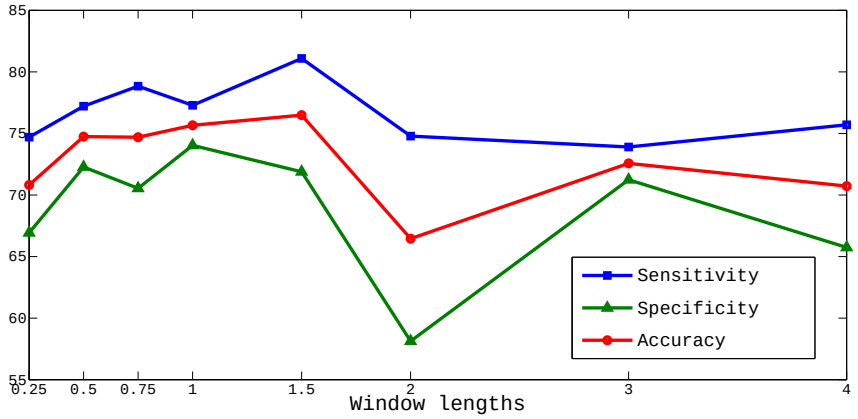


Figure 3.17: Sensitivity, specificity and accuracy for 12-feature, participant dependent LD classifiers. Results are displayed for window lengths of 0.25, 0.5, 0.75, 1, 1.5, 2, 3 and 4 seconds.

window lengths of 1 second. This may be due to longer window lengths, containing both normal and artefact data and shorter window lengths resulting in epochs which are too short to accurately reflect the characteristics of the signal.

Results from the mutual information calculations and LDC classifier indicate that the investigated feature set provides a good separation between artefact and background EEG classes for participant-dependent data. This confirms the validity of using a generalised head movement artefact class when detecting non-cerebral activity in the EEG induced by head movements. This finding is built upon in the next section, where more powerful classification techniques are used to perform participant-independent classification of head-movement artefacts.

3.5 Support vector machines

Linear discriminant classifiers perform best for Gaussian distributions. However, as is clear from the probability density functions displayed in Section 3.3, this is rarely the case for the features used in the artefact detection task. This provides a strong argument for use of a non-linear classifier in distinguishing between normal EEG and that contaminated by head movements. Support vector machines (SVM) is one such family of classifiers that tend to produce excellent classification performance for two-class problems; Gaussian kernel SVM classifiers have proven highly-effective in a number of other event detection systems in EEG (Temko et al., 2011b; Kelleher et al., 2010). SVMs benefit from the additional

advantage of being somewhat indifferent to "curse of dimensionality", thus allowing full advantage to be taken of the complete set of EEG features, rather than performing explicit feature reduction before the classification (Cristianini and Shawe-Taylor, 2000).

3.5.1 SVM theoretical background

The fundamental concept of the support vector machine is to transform a set of feature vectors $\mathbf{x}_j \in \mathbb{R}^d$ into a higher dimensional space and to find the optimal hyperplane in the space that can maximize the margin between classes. By transforming the data into a higher dimensional space, complex classification problems can be converted into more simple problems that can use linear discriminant functions. As the SVM is based only on those training patterns that are near the decision surface, classification calculation times are quite small, allowing for ease of implementation in a real-time system. Furthermore, as support vectors are a subset of the training data and contain all the information needed to define the classifier, SVMs are highly insensitive to the dimensionality of the feature space. This makes the SVM ideal for a system with a large number of features extracted from the EEG signal. The following paragraphs provide a brief description of support vector machines, in the context of two-class supervised learning.

Let a training feature vector $\mathbf{x}_j \in \mathbb{R}^N$ with associated labels $y_j \in \{1, -1\}$ belong to linearly separable classes, where the decision surface used to classify a pattern is the hyperplane $\mathbb{H}_{\nu}$. The problem of classifying a test vector $\mathbf{x}_k$ as belonging to one of two classes can be written as:

$$f(\mathbf{x}_k) = \mathbf{w} \cdot \mathbf{x}_k + b, \tag{3.38}$$

where $\mathbf{w} \in \mathbb{R}^N$. $\mathbb{H}_0$ is the region of vectors $\mathbf{x}$ which verify the equation $f(\mathbf{x}) = 0$. If $\mathbb{H}_1$ and $\mathbb{H}_{-1}$ are two hyperplanes parallel to $\mathbb{H}_0$, defined by $f(\mathbf{x}) = 1$ and $f(\mathbf{x}) = -1$, respectively, then the distance separating these two hyperplanes is:

$$d = \frac{2}{|| \mathbf{w} ||}. \tag{3.39}$$

The distance d, denoted the *margin*, is maximised so as to obtain a classifier boundary that is not overfit to the training data $\mathbf{x}_j$. To be correctly classified, the training vectors

should lie outside the margin or on the margin boundary and must satisfy:

$$\mathbf{w} \cdot \mathbf{x}_j + b \geq +1, \text{ for } y_j = +1, \tag{3.40}$$

$$\mathbf{w} \cdot \mathbf{x}_j + b \leq -1, \text{ for } y_j = -1. \tag{3.41}$$

This can be written more concisely as:

$$y_j(\mathbf{w} \cdot \mathbf{x}_j + b) \geq 1 \ \forall j. \tag{3.42}$$

The problem of finding the SVM classifying function $\mathbb{H}_0$ can be stated as:

$$\text{minimise } \frac{1}{2} \mid\mid \mathbf{w} \mid\mid^2, \text{ subject to } y_j(\mathbf{w} \cdot \mathbf{x}_j + b) \geq 1 \ \forall j. \tag{3.43}$$

As real-world biomedical data are often not linearly separable, with considerable overlap between classes, the decision boundary can be *softened* by introducing a *slack* positive variable ξ_j for each training vector. The conditions that the training vectors must satisfy, defined in Equations 3.40 and 3.41 can now be modified to include ξ_j such that:

$$\mathbf{w} \cdot \mathbf{x}_j + b \geq +1 - \xi_j, \text{ for } y_j = +1, \tag{3.44}$$

$$\mathbf{w} \cdot \mathbf{x}_j + b \leq -1 + \xi_j, \text{ for } y_j = -1. \tag{3.45}$$

The introduction of ξ_j is problematic in that the constraints in 3.44 and 3.45 will be met for all j if ξ_j is suitably large. To avoid trivial solutions, a regularisation constant C is introduced into the objective function which now becomes:

$$\text{minimise } \frac{1}{2} \mid\mid \mathbf{w} \mid\mid^2 + C \sum_{j=1}^{N} \xi_j, \ \text{ subject to } \ y_j(\mathbf{w} \cdot \mathbf{x}_j + b) \geq 1 \ \forall j. \tag{3.46}$$

The regularisation parameter C thus controls the degree of penalisation introduced by ξ_j, such that increasing C permits fewer training errors at the expense of reduced generalisation. The convex optimisation problem outlined by Equation 3.46 is solved with Lagrangian multipliers α_j. Only training patterns lying on the margin surface or within the margin have non-zero α_j; these are the support vectors. The classification process thus consists of assigning one of the two classes to a given input vector $\mathbf{x}_k$ of dimension N, such that:

$$f(\mathbf{x}_k) = \sum_{k=1}^{M} \alpha_k y_k \mathbf{x}_k \cdot \mathbf{x}_j + b, \tag{3.47}$$

where M is the number of support vectors and $M < N$.

In the case of classification problems that are non-linearly separable, the data is mapped to a higher dimensional feature space where the data are linearly separable, such that the classification function becomes:

$$f(\mathbf{x}_k) = \sum_{k=1}^{M} \alpha_k y_k \phi(\mathbf{x}_k) \cdot \phi(\mathbf{x}_j) + b, \tag{3.48}$$

where the dot product $\phi(\mathbf{x}_k) \cdot \phi(\mathbf{x}_j)$ can be replaced by a kernel function $\mathcal{K}$, such that:

$$\mathcal{K}(\mathbf{x}_j, \mathbf{x}_k) = \phi(\mathbf{x}_k) \cdot \phi(\mathbf{x}_j). \tag{3.49}$$

In this thesis, a Gaussian radial basis function (RBF) kernel is used, defined by:

$$\mathcal{K}(\mathbf{x}_j, \mathbf{x}_k) = e^{-\frac{1}{2\sigma^2}|\mathbf{x}_k - \mathbf{x}_j|^2}. \tag{3.50}$$

In practice, it is not necessary to know the mapping function ϕ explicitly, thanks to a mathematical property known as the *kernel trick* (Schölkopf, 2001). In this context, only the kernel $\mathcal{K}(\mathbf{x}_j, \mathbf{x}_k)$ must be known, and can be interpreted as a non-linear similarity measure between the two data points. The SVM classifier is thus:

$$y(\mathbf{x}_k) = sgn\left(\sum_{l=1}^{M} \alpha_l y_l K(\mathbf{x}_k, \mathbf{x}_l) + b \right), \tag{3.51}$$

where M is the number of support vectors.

The output scores of the SVM can be converted to a probabilistic estimation of class by means of a sigmoid function:

$$P(Art|x) \approx P_{A,B}(f) = \frac{1}{1 + exp(Af + B)}, \tag{3.52}$$

where $P(Art|x)$ is the probability that an epoch contains artefact activity, $f = f(x)$ is the distance to the separating hyperplane (i.e. the output of the SVM classifier), and A and B are the parameters of the sigmoid function estimated on the training dataset (Platt, 1999). This probability output $P(Art|x)$ is compared to a threshold value θ, resulting in a binary decision; 1 indicating artefact activity and 0 indicating normal, background EEG. Thus, by specifying the threshold θ, the classifier can be made to operate in a chosen region of the ROC curve (Section 2.9.4).

3.5.2 SVM model selection

The head-movement artefact classification task was implemented with Gaussian RBF-kernel SVMs and repeated for each of the window lengths investigated with LDCs. An outline of the entire SVM classification architecture is displayed in Figure 3.18. In the classification task outlined here, SVMs were trained and tested on participant independent data sets; i.e. the SVM classifier used to detect artefacts in a participant's EEG recording was not trained on any data from that participant.

Leave-one-out cross-validation (LOO-CV) is used to evaluate the participant independent SVM classifier performance. In this manner, all but one of the participants' data is used for training and the remaining participants' data is used for testing. This procedure is repeated until each participant has been a test subject and the mean result is reported. The leave-one-out method is known to be an almost unbiased estimation of the true generalization error; i.e., the performance reported with the leave-one-out method is the most similar to the performance this system would show on an unseen test dataset of infinite length once it is trained on all available data (Vapnik, 2006). As deploying such an EEG system outside of academia would likely mean performing artefact detection on an unseen participant, the process of training the SVMs on different participants than those on which it will be tested (and repeating for each participant), the performance that would likely be encountered in a real-world deployment is approximated. trained in this thesis best account for variations between participants, The SVMs implemented in this work are based on the LIBSVM library (Chang and Lin, 2011).

In practice, the regularisation parameter C and the kernel variance term σ^2 must be specified for an RBF-kernel SVM. In this thesis, these parameters are chosen by performing 5-fold cross validation on the training data (Kohavi, 1995); the data is randomly split into five equal parts, four-fifths of the data is used for training the model, and the remaining fifth is used to evaluate the performance. This process is repeated five times and the best parameter set is chosen.

3.5.3 SVM performance

ROC areas for the participant independent SVMs are displayed in Figure 3.19. Boxplots summarise the ROC areas across participants here for the eight window lengths investigated. Boxplots display differences between populations without making any assumptions

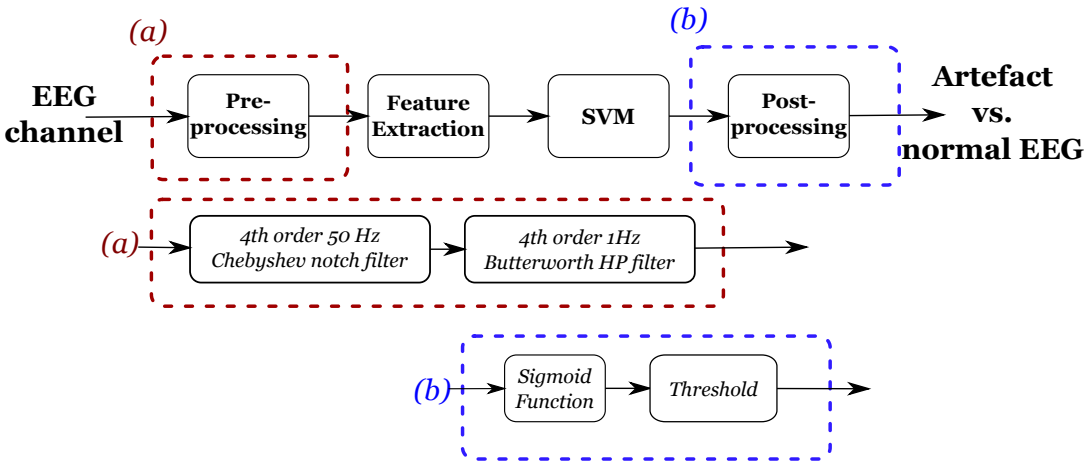


Figure 3.18: Head-movement artefact detection system based on support vector machines classifier. Classification is performed on each EEG channel separately.

about the underlying statistical distribution, i.e. they are non-parametric. The spacings between the different parts of the box indicate the degree of spread and skewness in the data, and identify outliers. Boxplots summarise the data with five numbers: the smallest observation (sample minimum), largest observation (sample maximum), lower quartile (Q1), median (Q2), upper quartile (Q3). Sample minima and maxima are indicated by the horizontal black lines at the end of the whiskers. Lower and upper quartiles are indicated by the horizontal blue sides of the boxes. Sample median is indicated by the horizontal red lines.

Classification results for the generalised head-movement artefact validated the feasibility of using of a generalised head-movement class in detecting artefacts using participant independent data sets. ROC area was highest for a window length of 0.75 seconds at 83 %; average ROC area was 80.2 % across window lengths (σ = 2.7%). However, the differences between ROC areas across window lengths was small when considering the variation in ROC areas between participants.

3.6 Summary

In this chapter, artefact detection for ambulatory EEG was introduced. A statistical machine learning approach was taken to the problem of identifying EEG sections that are contaminated by electrical activity arising from non-cerebral sources. Thus, a comprehensive feature set (comprising time domain, frequency domain and information theoretic

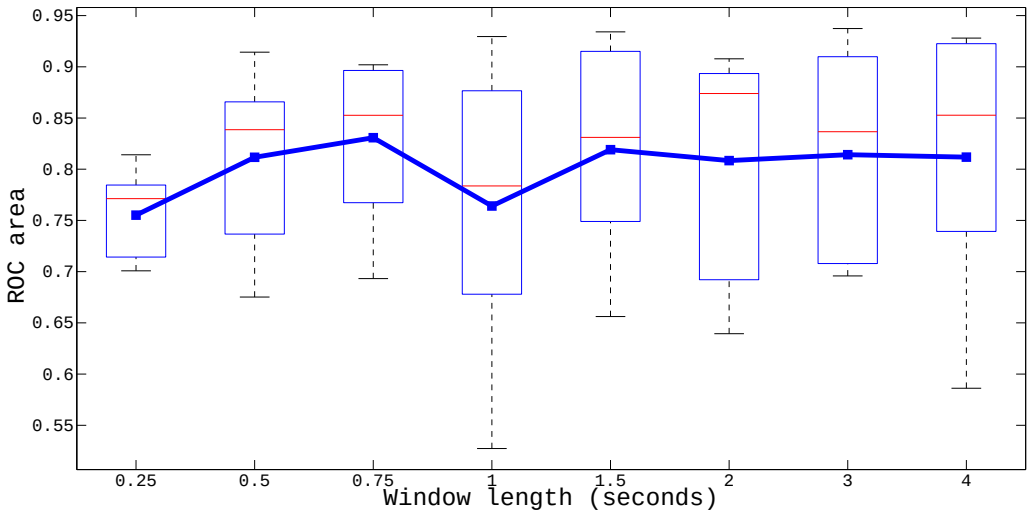


Figure 3.19: ROC areas of participant independent SVM classifiers for a range of window lengths.

features) was compiled and investigated for its ability to provide separation between normal EEG and EEG that has been contaminated by artefact following head-movement. Inspection of the features' probability density functions and mutual information with the class labels indicated their usefulness in the generalised head-movement artefact detection task. This concept of generalised head-movement artefact detection was then validated using a reduced feature set and a linear discriminant classifier; classification accuracies of between 65 % and 70% were recorded for each of the window lengths investigated with a window length of 1.5 seconds showing the best performance. With the feasibility of detecting head-movement artefacts with a single classifier confirmed, classification using a support vector machines classifier was performed utilising the full feature set. Mean ROC areas of between 75 % and 83 % were observed, this time on the more difficult task of participant independent artefact detection. A window length of 0.75 seconds resulted in the highest mean ROC area of 83 % ; however, given the variation in classification performance across participants, a definitively best window length was not chosen. The SVM classifier trained and tested in this chapter is thus suitable for artefact annotation of EEG recorded in an ambulatory environment, and to the author's knowledge is the first of its kind. Additionally, this artefact detector could be incorporated into an automated neurological event detection system at either the pre-processing (via artefact rejection) or post-processing stage (via classifier fusion).

Chapter 4

Detecting head-movement artefacts with gyroscopes

4.1 Introduction

It was shown in Chapter 3 that as head movements often result in EEG artefacts (and certain head-movements may be more likely than others to result in artefact), it follows that information regarding movements of the head could lead to insights into the subsequent production of artefact in the EEG. Gyroscopes, devices that measure angular rotation, offer a means of accurately detecting head movements. Recent advances in the miniaturisation of gyroscopes have resulted in their inclusion in the commercially available Emotiv EPOC EEG headset (Emotiv EPOC headset). This chapter investigates the use of the gyroscopes built into the Emotiv EPOC headset in providing information on head movement and subsequently using this information to detect resultant contamination of the EEG signals. A journal paper (O'Regan et al., 2013a) and an international peer-reviewed conference paper (O'Regan et al., 2010b) have arisen from the work carried out in this chapter.

4.2 Gyroscopes

The advent of miniaturized sensing technology has paved the way for inconspicuous body-worn sensors that allow the collection and storage of data measuring the different aspects

of human movement. Accordingly, a vast body of literature has grown in conjunction with this, outlining ways in which advanced signal processing and statistical pattern recognition techniques can be applied to signals generated by inertial sensors such as gyroscopes and accelerometers in order to provide information on various forms of body movements and physical activities. Gyroscopes have been effectively deployed to capture motion for a wide range of biomedical applications, e.g. Bourke and Lyons (2008) used bi-axial gyroscopes for fall-detection. Several papers have outlined the use of gyroscopes in conjunction with accelerometers for fall detection in the elderly (Greene et al., 2010), gait analysis during walking and running (McGrath et al., 2012) and gait assessment of patients with Parkinson's disease (Salarian et al., 2004).

The Emotiv EPOC headset is equipped with two microelectromechanical systems (MEMS) gyroscopes located on the printed circuit board at the back of the device. These gyroscopes were initially included in the EPOC for attitudinal control in BCI applications. However, this set-up offers an inexpensive method of utilising inertial sensors to gain insight into head-movements and consequently into the head-movement artefact detection task. Internally, the MEMS gyroscopes use lithographically constructed vibrating wheels. When movement occurs, a wheel is driven to rotate a fraction of a full turn about its axis. The tilt of the wheel is then measured to produce a signal related to the rate of rotation (Bernstein, 2003). The gyroscopes in the Emotiv EPOC provide measurements of angular velocity within two planes. In this work, movement within these two planes are referred to as x-direction and y-direction movement; x-direction movement refers to lateral rotation around the neck axis and y-direction movement indicates vertical rotation around the axis joining the ears. These axes are illustrated in Figure 4.1.

Visual inspection of the EEG and gyroscope signals outlined in Figure 4.2 shows clear correlation between each of the two gyroscope signals and the EEG trace at times of head-movements. However, as x-direction and y-direction gyroscope signals each detect movement about one axis only, each signal is appropriate for the detection of certain types of movements. Thus, the y-direction gyroscope signal adequately highlights periods of EEG artefacts caused by nodding the head, but is considerably less successful in detecting the EEG artefacts brought about by shaking the head. Conversely, the x-direction gyroscope signal barely registers the nodding of the head, but comprehensively flags shaking of the head in a manner that is clearly correlated to sections of head-movement artefact. To combine information from both gyroscopes into a single, composite gyroscope signal, absolute angular velocity $\alpha(n)$ is created by taking the square-root of the sum of the squared

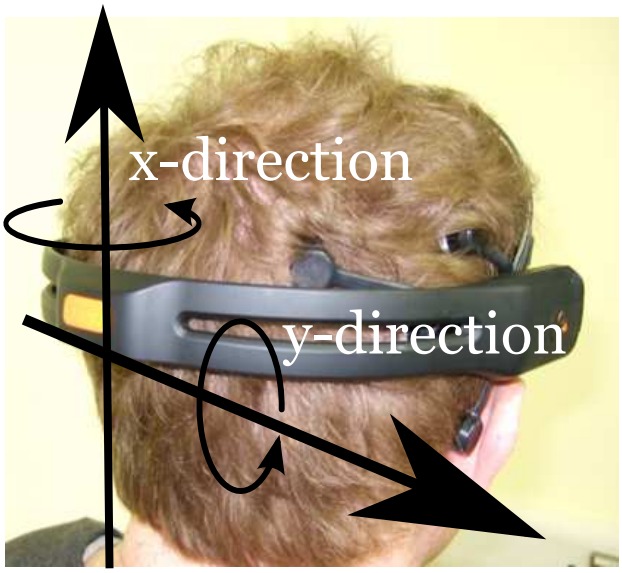


Figure 4.1: Emotiv EPOC headset: Gyroscope axes.

x- and y- direction gyroscope signals:

$$\alpha(n) = \sqrt{gyro_x(n)^2 + gyro_y(n)^2}. \tag{4.1}$$

This value quantifies the angular velocity that takes place in either direction, giving a measure of overall head-movement and indicating whether or not head-movement has occurred. As there may be an unknown delay between the head-movement as detected by the gyroscope, and the appearance of the resultant artefact on the EEG trace, two different moving average (MA) filters are applied to $\alpha(n)$ to produce two additional derivative gyroscope signals. The length of the MA filters was determined based on observation. A 10-point MA filter was applied to the absolute angular velocity signal $\alpha(n)$ as described in Equation 4.2.

$$MA_{10} = \frac{1}{10}\sum_{k=0}^{9}\alpha(n-k). \tag{4.2}$$

Similarly, in Equation 4.3 a 50-point MA filter was applied to $\alpha(n)$:

$$MA_{50} = \frac{1}{50}\sum_{k=0}^{49}\alpha(n-k). \tag{4.3}$$

An example of the five gyroscope signals are displayed in Figure 4.3, for thirty seconds of data, showing nodding and shaking of the head in addition to sections of background EEG activity where no head-movement is present. Similarly, Figures 4.4 and 4.5 show EEG and gyroscope signals for artefacts arising from raising and lowering of the eyebrows

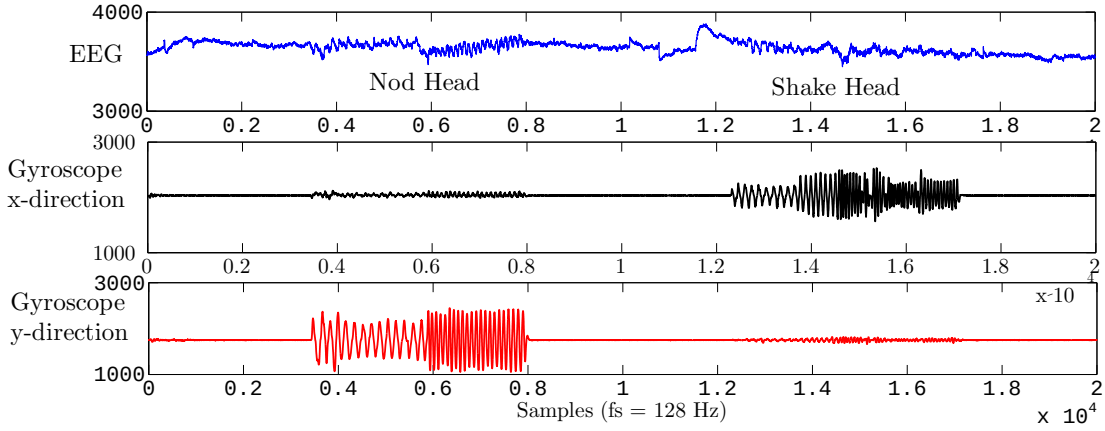


Figure 4.2: EEG channel F8 (blue) and gyroscope x-direction (black) and y-direction (red) signals for artefacts caused by raising nodding and shaking head.

and from clenching the jaw, respectively.

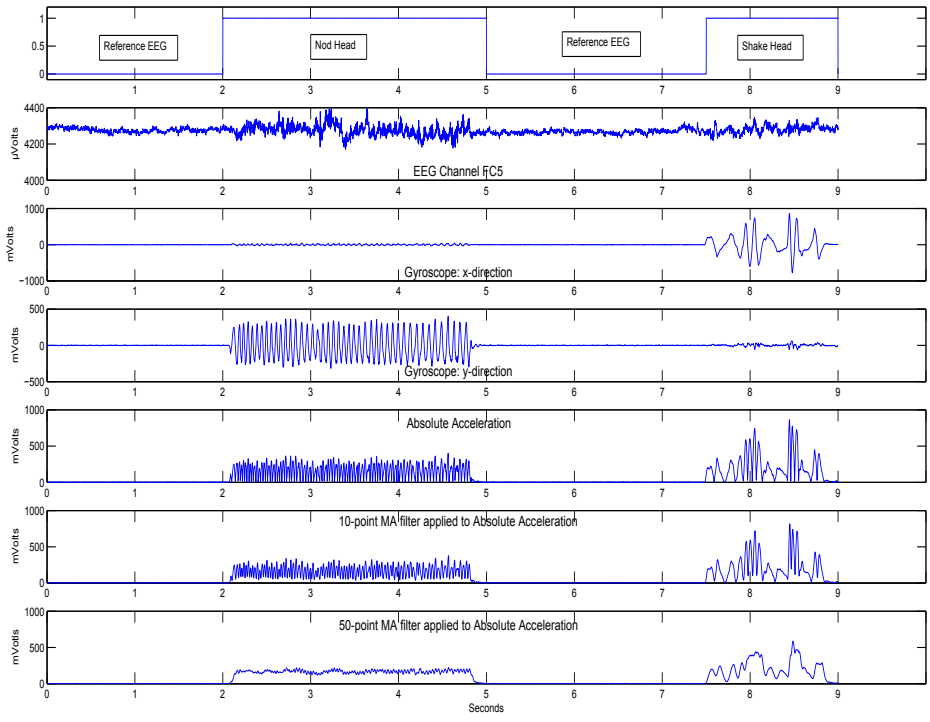


Figure 4.3: EEG channel FC5 and all 5 gyroscope channels for artefacts caused by nodding and shaking of the head, in addition to sections of background EEG activity.

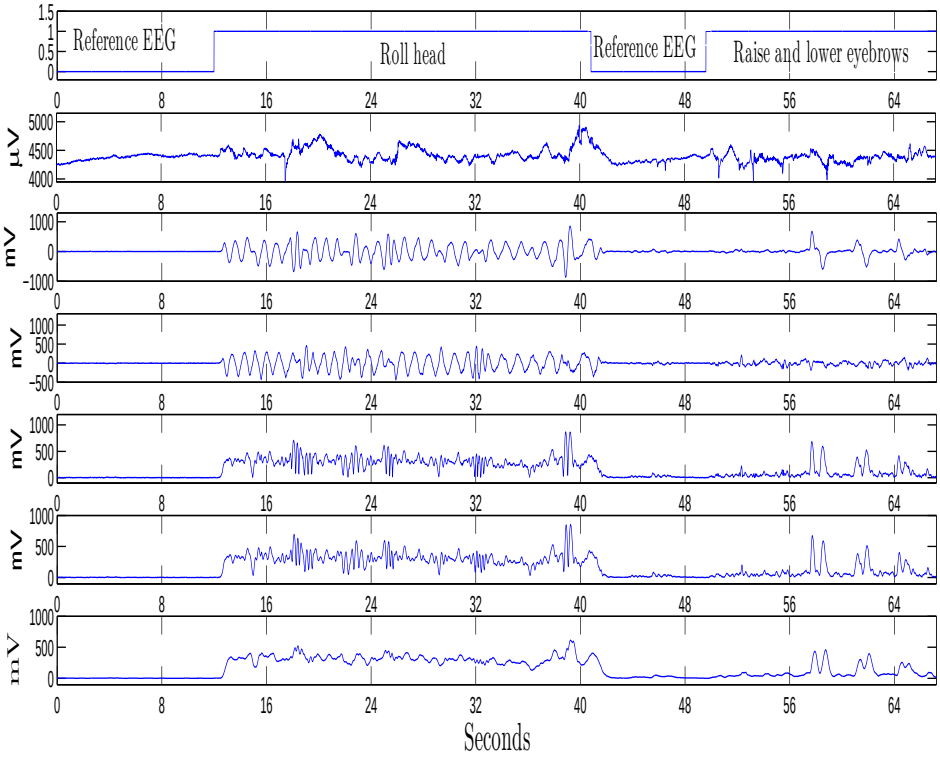


Figure 4.4: EEG electrode F3 and all 5 gyroscope channels for artefacts caused by raising eyebrows and roll head.

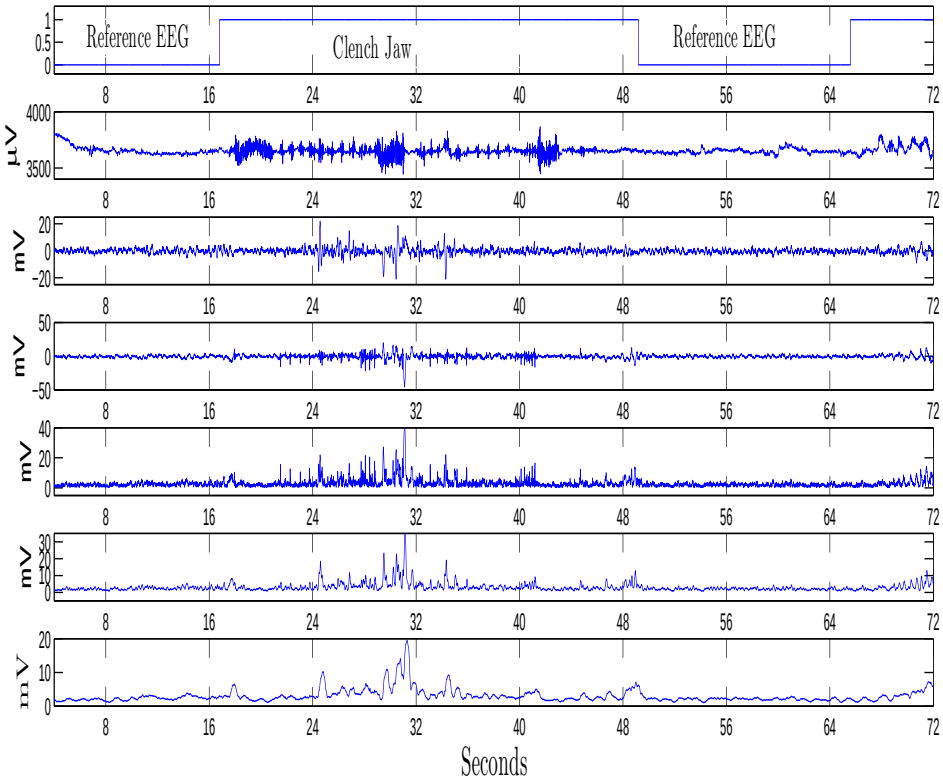
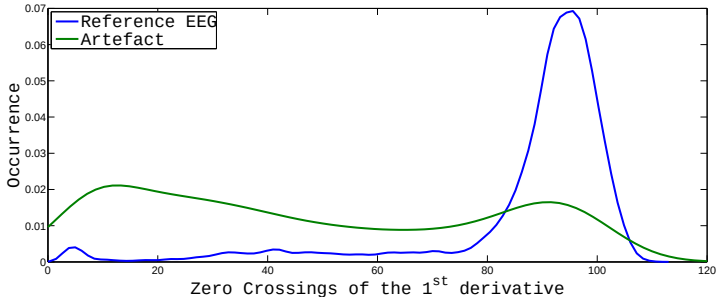


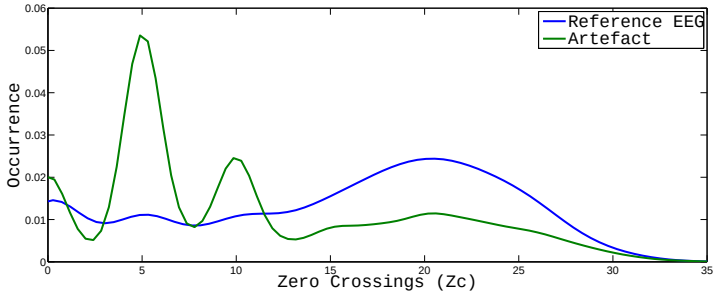
Figure 4.5: EEG electrode F7 and all 5 gyroscope channels for artefacts caused by clenched jaw.

4.3 Gyroscope features

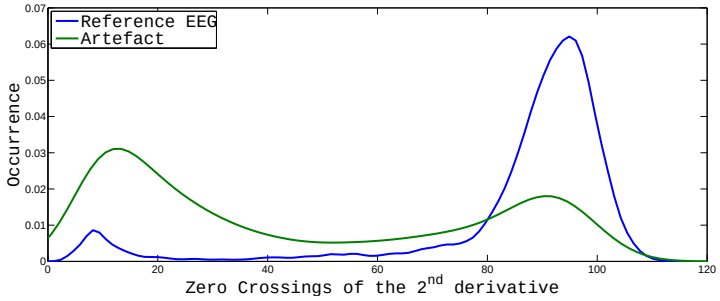
As discussed in Chapter 3, in statistical pattern recognition, features are quantitative descriptors of a signal whose purpose is to show significant changes in value between classes. In this experiment, periods where head-movements lead to EEG artefact are known as they were carefully annotated when the data was recorded (Section 2.9.1). The classification goal is thus to find a discriminant function that separates between features extracted from the gyroscope signals so that those features corresponding to times when head-movement artefact is produced lie on one side of the function, and features extracted from gyroscope signals at times when the head remains still lie to other side of the function. In Figures 4.3, 4.4 and 4.5 there are clear visual correlations between artefact in the EEG and fast, high-amplitude deflections in the gyroscope signal. To quantitatively capture this, the five gyroscope signals (x, y, α, MA_{10} and MA_{50}) were segmented into overlapping epochs and a number of time domain features as well as one feature from the frequency domain (total power) were extracted from each gyroscope signal for each epoch. Gyroscope features were chosen to reflect the fact that when head-movement occurs, the gyroscope signal undergoes relatively large deflections that oscillate with changing direction. In particular, it was observed that larger and more abrupt head-movements were inclined to produce artefacts in the EEG. Unlike the EEG signal however, the gyroscope signals exhibited a much more limited frequency range with less complex morphologies observed; consequently only one frequency-based feature was inclued. These features were extracted from each of the five gyroscope channels, providing a total of 80 gyroscope features extracted for each epoch. To the author's knowledge, the work carried out in this chapter is the first to utilise gyroscopes to detect the presence of head-movement artefact in EEG. It was not clear what window length was most appropriate for this classification task; a range of window lengths were thus investigated, with segmentation performed for window lengths of 0.25, 0.5, 0.75, 1, 1.5, 2, 3, and 4 seconds.



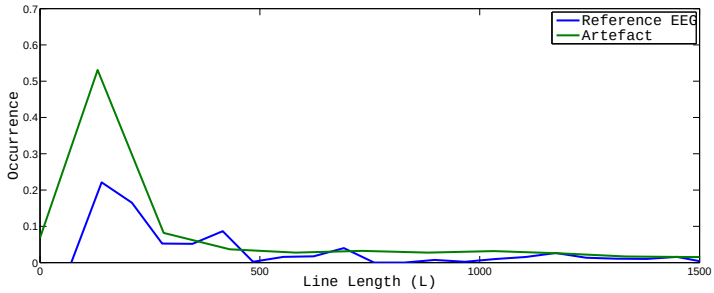
(a) Zero crossings of the 1^{st} derivative of y-direction gyroscope signal.



(b) Zero crossings of the 1^{st} derivative of x-direction gyroscope signal.

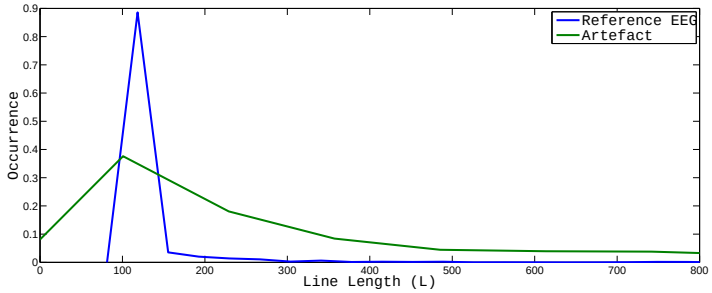


(c) Zero crossings of the 2^{nd} derivative of MA-10 applied to the absolute angular velocity gyroscope signal $\alpha(n)$.

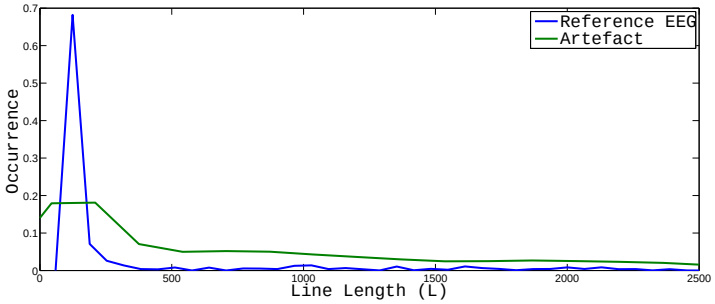


(d) Line length of x-direction gyroscope signal.

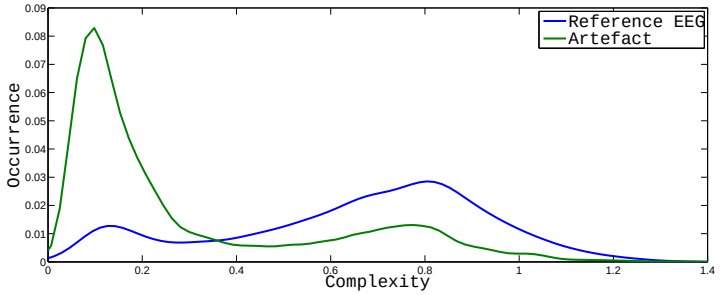
Figure 4.6: Probability density functions of a selection of gyroscope features corresponding to sections of EEG where head-movement artefact was registered and gyroscope features corresponding to background EEG. The pdfs displayed describe features extracted for window lengths of 1 second.



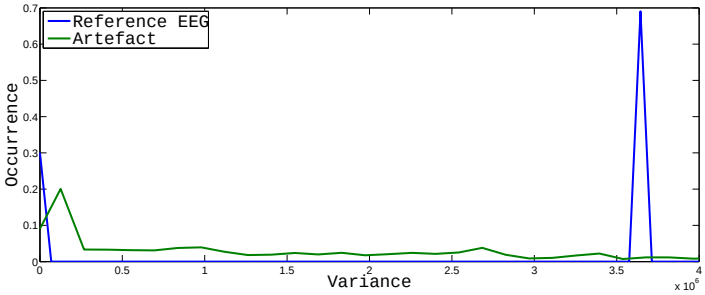
(a) Line length of y-direction gyroscope signal.



(b) Line length of the absolute angular velocity gyroscope signal $\alpha(n)$.



(c) Complexity of the y-direction gyroscope signal.



(d) Variance of the 1^{st} derivative of y-direction gyroscope signal.

Figure 4.7: Probability density functions of a selection of gyroscope features corresponding to sections of EEG where head-movement artefact was registered and gyroscope features corresponding to background EEG. The pdfs displayed describe features extracted for window lengths of 1 second.

Table 4.1: List of gyroscope features for head-movement artefact detection.

Time Domain Features
• Variance
• Zero Crossings
• Nonlinear Energy (NLE)
• Line Length/ Curve Length
• No. of inactive samples
• Activity (1st Hjorth Parameter)
• Mobility (2nd Hjorth Parameter)
• Complexity (3rd Hjorth Parameter)
• RMS Amplitude
• Kurtosis
• Skewness
• Zero Crossings of first derivative
• Zero Crossings of second derivative
• Variance of first derivative
• Variance of second derivative
Frequency Domain Features
• Total Power

The complete list of gyroscope features is displayed in Table 4.1 and described in the following paragraphs. The discriminating potential of individual gyroscope features can be visualised by comparing the probability density density functions (pdf) of features from two classes (gyroscope signals corresponding to artefact EEG and gyroscope signals corresponding to normal EEG). Probability density functions for several gyroscope features are illustrated in Figures 4.6 and 4.6. The following paragraphs provide an interpretation of these pdfs and are included here as an illustration of the reasoning behind the choice of certain features.

No. of zero crossings of the gyroscope signal and its derivatives

The number of zero crossings (Zc) is the number of times within an epoch that the gyroscope signal crosses the x-axis. The number of zero crossings of the 1st derivative of the gyroscope corresponds to the number of local maxima and minima of the gyroscope signal. The number of zero crossings of 2nd derivative corresponds to the number of times that the 2nd derivative of the EEG signal crosses the x-axis within an epoch. During periods when the participant remains still, the gyroscope signal is centred about the x-axis with small deviations representing minute movements of the headset and consequently crosses the x-axis frequently. Due to the large deflections in the gyroscope signals when head-movement occurs, the number of zero crossings of the gyroscope and its derivatives should be smaller in sections of head-movement.

Probability density functions for zero crossings of the 1^{st} derivative are displayed in Figure 4.6a. The reference EEG and artefact signals are bimodal; the reference EEG has a clear, dominant peak centred at approximately 90 crossings per epoch. The peaks of the distribution corresponding to artefact are more equally sized, with a wide range of values observed. These trends can be interpreted by considering that this feature measures the rate of change of the gyroscope signal in a given direction. When little head movement is observed (corresponding generally to normal EEG), the gyroscope signal will be low-amplitude, deviating slightly but frequently with each minor movement of the headset. Consequently, in a given epoch, there should be a relatively large number of zero crossings. During head-movement the situation is quite different. When the head undergoes large changes in direction, a high amplitude gyroscope signal results. The derivative of this signal is then positive when the signal is increasing and negative when the signal is decreasing. Due to the rate of change of these high amplitude signals, the derivative of the y-direction gyroscope will thus have fewer zero crossings for these movements, a fact that is captured in the larger peak of the distribution. However, it should be noted that the y-direction gyroscope signal can only record head movements about the axis joining the ears; any movements about the axis defined by the neck will be undetected and appear as if no head-movement has occurred. Accordingly, the second peak of the distribution corresponds to sections of head-movement which takes place about an axis that is unobserved by the y-direction gyroscope signal (e.g. shaking the head). With some notable differences, the pdf of zero crossings of the 1^{st} derivative of the x-direction gyroscope signal follows a broadly similar pattern of discrimination between classes as that observed for the y-direction movement and is illustrated in Figure 4.6b. Once again, features corresponding to background EEG tend to exhibit higher numbers of zero crossings than features corresponding to artefactual EEG. However, the distribution is multimodal for the gyroscope features corresponding to artefact EEG, with individual peaks describing different types of head-movements, with different characteristic angular velocities.

Figure 4.6c displays probability density functions for zero crossings of the 2^{nd} derivative of $\alpha(n)$ with a 10-point moving average filter applied. Like Figure 4.6a, the distributions are bimodal; the dominant peak of the signal corresponding to normal EEG located separately to a well-defined peak of the signal corresponding to artefact EEG. Interpretation of the discriminative peaks is analogous to that of Figures 4.6a and 4.6b, with acceleration of the gyroscope signal replacing rate of change of the gyroscope signal.

Nonlinear Energy (NLE)

Non-linear energy, given by Equation 3.10 was included as an EEG feature in Section 3.3. Gyroscope signals during periods of head-movement typically contain more energy than periods without movement, and consequently should exhibit higher NLE.

Line Length/ Curve Length

Line length, given by Equation 3.1 describes the running sum of distances between consecutive points within a sliding window, and gives a measure of signal complexity. Accordingly, gyroscope signals corresponding to movements, where there are big amplitude jumps between samples, are expected to have considerably higher line length than gyroscope signals corresponding to no head-movements.

Line length of the x-direction, y-direction and absolute angular velocity gyroscope signals are displayed in Figures 4.6d, 4.7a and 4.7b, respectively. While there is considerable overlap between distributions, in general, gyroscope features corresponding to normal EEG tend to be predominantly of values between 100 and 200. In comparison to this, the gyroscope features corresponding to EEG artefact is more broadly distributed. That is, gyroscope signals while no head-movements take place is quite predictable in terms of line length. However, during head-movements line length tends to be quite varied.

Maximum value of autocorrelation

The maximum value of the autocorrelation $R_{ss}(\mathbf{x}_j)$ of a gyroscope signal was used by Tunçel et al. (2009) to classify human leg movements. The autocorrelation is defined as:

$$R_{ss}(\mathbf{x}_j) = \frac{1}{N_s - \Delta} \sum_{i=0}^{N_s-\Delta-1} \left(s_j(i) - \mu_j\right)\left(s_j(i-\Delta) - \mu_j\right), \tag{4.4}$$

where $\Delta = 0, 1, ..., N_s - 1$. Repetitive movements such as those produced during shaking of the head have higher maximum autocorrelation values than irregular movements or periods of rest.

No. of inactive samples

The number of inactive samples within an epoch $\mathbf{x}_j$ is defined as the number of samples for which there is very little change in the gyroscope amplitude. This was calculated by applying a threshold of 0.01 to the absolute value of the derivative of the gyroscope sig-

nal. Gyroscope sections without head-movements should therefore have higher numbers of inactive samples.

Hjorth parameters

The Hjorth parameters (activity/variance, mobility and complexity) given by Equations 3.4, 3.6 and 3.7 are commonly used in EEG signal processing. These features are utilised here for gyroscope signals. Owing to the fact that a large range of head-movements are examined in this experiment, it is anticipated that the variance of gyroscope signals corresponding to artefactual sections of EEG should on average be greater than the variance of those corresponding to normal EEG. The probability density function of the complexity of the y-direction gyroscope signal is displayed in Figure 4.7c. Clear discrimination can be observed between the distributions; gyroscope epochs corresponding to artefact EEG exhibiting less complexity (due to more regular gyroscope signals) than those of gyroscope epochs corresponding to normal EEG. As was observed for other features, the range of values is not confined to those within the dominant peak but manifests as a wide spread of values. This phenomenon is likely explained by movements that are not measurable about the y-direction axis.

Root mean square amplitude

The root mean square (RMS) amplitude, or quadratic mean, is a statistical measure of the magnitude of a time varying quantity and given by Equation 3.2. As gyroscope signals corresponding to artefact EEG are often high-energy, high-amplitude signals, the RMS amplitude aims to capture this trait.

Skewness

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable. Skewness has been used to classify human leg movements with gyroscopes in Tunçel et al. (2009) and is given in Equation 3.8. Owing to the variety of different head-movements, the distribution of gyroscope signals corresponding to sections of EEG where artefacts are produced are hypothesized to have higher skew than those corresponding to the condition of no movement.

Kurtosis

Kurtosis, often referred to as the fourth central moment, is a measure of the "peakedness" of a probability density function and is defined in Equation 3.9.

Variance of first and second derivatives

The variance of the first ($\sigma_{\Delta j}$) and second ($\sigma_{\Delta 2 j}$) derivatives were used by Thomas et al. (2010) for EEG seizure detection. They are included here as gyroscope features in the artefact classification task as they were found to provide reasonable separation between gyroscope signals during rest and signals during movements.

Variance of the 1^{st} derivative of the y-direction gyroscope signal is displayed in Figure 4.7d. The variance of the gyroscope features corresponding to artefactual EEG is generally lower than that of features corresponding to normal EEG. In essence, this feature is another example of a trend visible in many of the illustrated features; namely, head-movement tends to bring a degree of order to the gyroscope signal.

Total Power

The total power refers to the sum of power in all bins of the power spectral density (PSD) between 0 and 12 Hz and is given by Equation 3.17. This feature is common in gyroscope classifiers and refers to the total power in the frequency range of typically occurring gyroscope signals. Periods of movement usually exhibit higher total power than inactive sections.

To summarise, clear discrimination between pdfs from gyroscope signals corresponding to normal and artefactual EEG are visible in varying degrees in Figures 4.6 and 4.7. Caution should be exercised however in concluding that a feature with strongly overlapping pdfs indicates the uselessness of that feature. As explained in Chapter 3, this is not the case; multiple features that are independently non-discriminative may provide adequate separation between classes when combined.

4.4 Feature reduction and linear discriminant classification

As shown in Chapter 3, the simplicity of linear discriminant classifiers provide a good starting point from which to begin an investigation into an uncharted classification task. Therefore, as an initial experiment to investigate the feasibility of using gyroscope features

to detect head-movement artefact in the EEG, a number of linear discriminant classifiers (LDC) were trained and tested on a subset of the features outlined in Section 4.3. As discussed in Section 3.4.1, with such a large feature set, it is likely that the information captured by some features may duplicate the information contained within other features and it may be beneficial to reduce the number of features by selecting a smaller feature subset and so avoid the "curse of dimensionality" (Abu-Mostafa et al., 2012). As was the case for the EEG linear discriminant classifier in Chapter 3, a multivariate filter technique, mutual information evaluation function (MIEF), performed the dual role of estimating the usefulness of the individual features and of reducing the feature set for the linear discriminant classifiers.

4.4.1 Mutual information evaluation function

The issue of feature ranking is discussed in some detail in Section 3.4.1, where it was shown that the mutual information evaluation function outlined by Al-Ani and Deriche (2002) was a suitable means of ranking features in terms of predicted utility for a classification task. Consequently, MIEF was employed to rank the gyroscope features in terms of usefulness for the head-movement classification task. MIEF performed feature ranking by calculating the mutual information between a gyroscope feature and the EEG class labels, the mutual information between the gyroscope feature and the previously selected features, and the joint mutual information between the gyroscope feature and previously selected features and the EEG class labels. In this manner, MIEF calculates the information gain brought to the classification task by each additional gyroscope feature. MIEF was

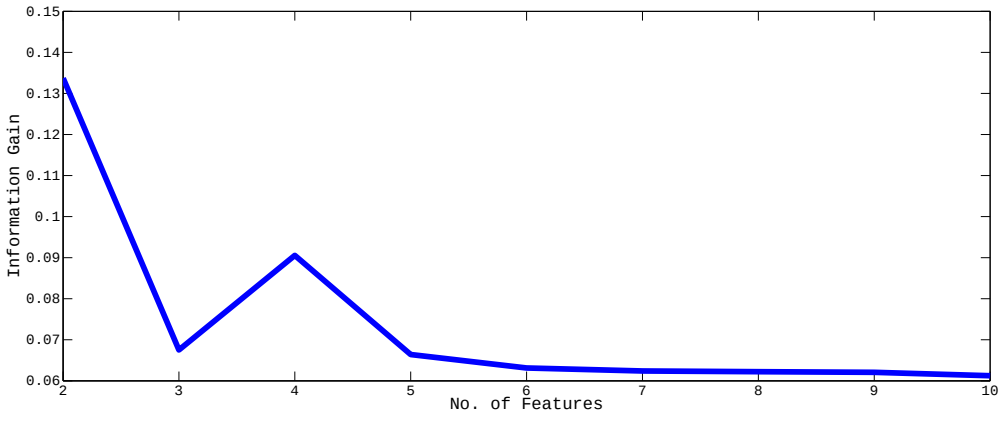


Figure 4.8: Information gain as function of the number of features used, as ranked by the mutual information evaluation function for a window length of 1 second.

performed for window lengths of 0.25, 0.5, 0.75, 1, 1.5, 2, 3, and 4 seconds. The best-performing features are displayed in Table 4.2, where it can be seen that certain features performed consistently well across window lengths. The best performing features were those associated with gyroscope y-direction movement and the two, filtered $\alpha(n)$ signals, MA-10 and MA-50. That is to say, features derived from up-down head movement and absolute head-movement were most useful in indicating if head-movement artefacts were produced in the EEG recording. In particular, features that captured the amplitude (line length) or rate of oscillation (zero crossings of the 1^{st} and 2^{nd} derivatives) of the gyroscope signals were ranked highly. Across features, there was significant overlap in information, indicating that the primary advantages of using gyroscopic features occurs with the addition of the first few features and information gain tapers off thereafter. This is seen in Figure 4.8, where the information gain flattens with the addtion of the sixth feature. Note also, in Figure 4.8 that due to the fact that the cost function of MIEF is not monotonically decreasing, the addition of a feature may result in increased information gain; thus, the spike in information gain with the addition of the fourth feature.

Table 4.2: Best-performing features using MIEF for window lengths 0.25 to 4 seconds.

0.25	**0.5**	**0.75**	**1**	**1.5**	**2**	**3**	**4**
20	20	20	61	20	61	68	77
32	61	61	68	77	68	61	68
77	68	68	20	28	20	76	76
28	28	28	28	68	30	77	61
36	22	22	77	36	28	20	30

Feature 20	y-direction gyro, line length (L)
Feature 28	y-direction gyro, Zero Crossings of the 1st derivative
Feature 36	absolute acceleration, line length (L)
Feature 61	MA-10, Zero Crossings of the 2nd derivative
Feature 68	MA-50, line length (L)
Feature 77	MA-50, Zero Crossings of the 2nd derivative

4.4.2 Linear discriminant classification

Linear discriminant classifiers find a linear combination of features that characterise or separate the two classes presented in the head-movement artefact detection classification task. A detailed account of the underlying theory associated with linear discriminant analysis is outlined in Section 3.4.3. The system architecture for the gyroscope linear discriminant classifier is illustrated in Figure 4.9.

LDC model selection

In this experiment, participant-dependent linear discriminant classifiers were trained using 5-fold cross validation; training was performed on a random four fifths of the data, and

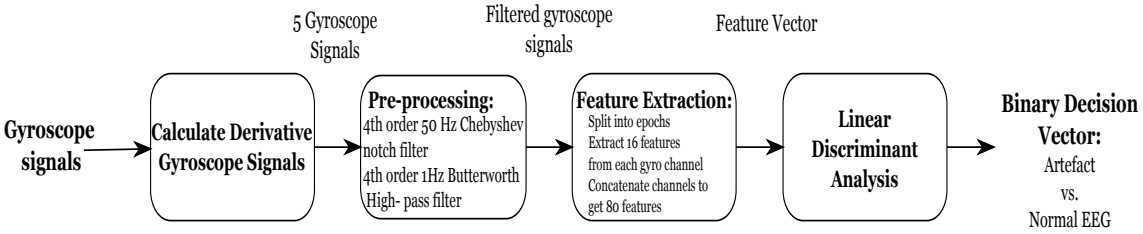


Figure 4.9: System architecture for gyroscope linear discriminant analysis head-movement artefact detector.

tested on the remaining fifth of data points and this process was repeated five times. This was performed for each of the investigated window lengths: 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4 seconds. For each window length, a feature set of the best-performing features were identified using MIEF as displayed in Table 4.2. Feature subsets of the top 3, 4, and 5 highest-ranking features from the MIEF algorithm were chosen for the classification task.

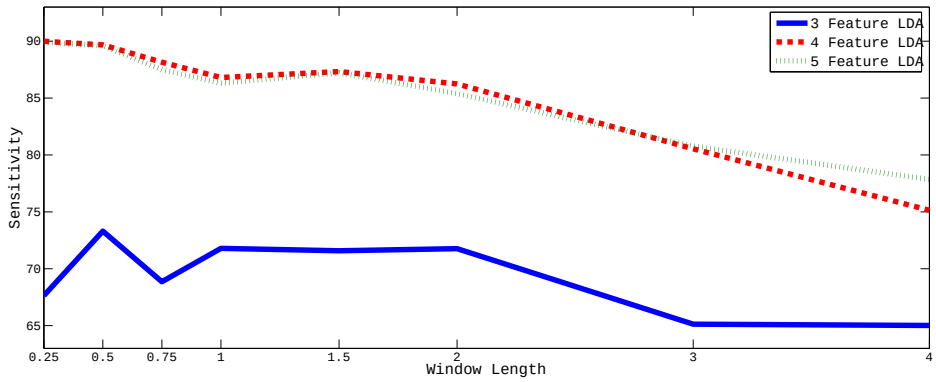
LDC performance

The performance of the classifiers was evaluated by sensitivity, specificity and accuracy metrics; the reported results are the mean of the five-fold CV process. The LD classification results from the 3-, 4-, and 5-feature LD classifiers are displayed in the form of sensitivity, specificity and accuracy in Figures 4.10a, 4.10b and 4.10c, respectively.

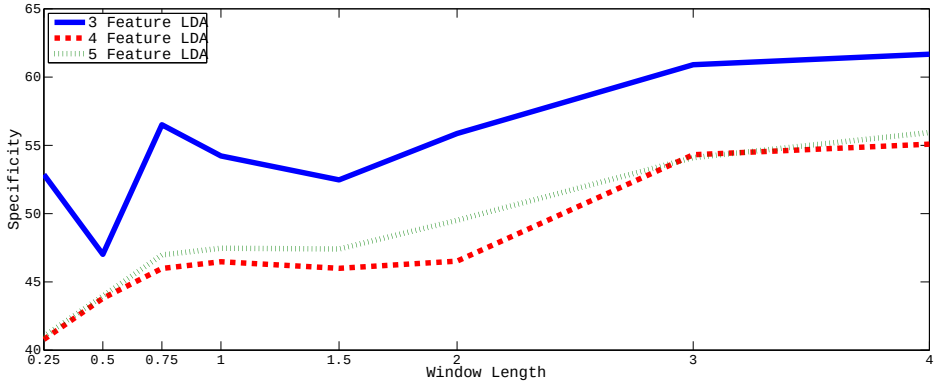
Sensitivity values of the 4- and 5-feature linear discriminant classifiers were considerably higher than those of the 3-feature classifiers (Figure 4.10a). Conversely, specificities of the 3-feature LDC outperformed its 4- and 5-feature counterparts (Figure 4.10b). The addition of features allows the accurate detection of more artefact epochs, at the cost of reduced specificity. Figure 4.10c displays the accuracy of the classifiers, indicating that despite the reduction in specificity brought on by the addition of the fourth and fifth features, it is outweighed by the corresponding improvement in sensitivities. Overall, results from the LD classifiers indicate that using the gyroscope features to separate between normal EEG and artefacts induced by head movement is possible and warrants further investigation.

4.5 Support vector machines

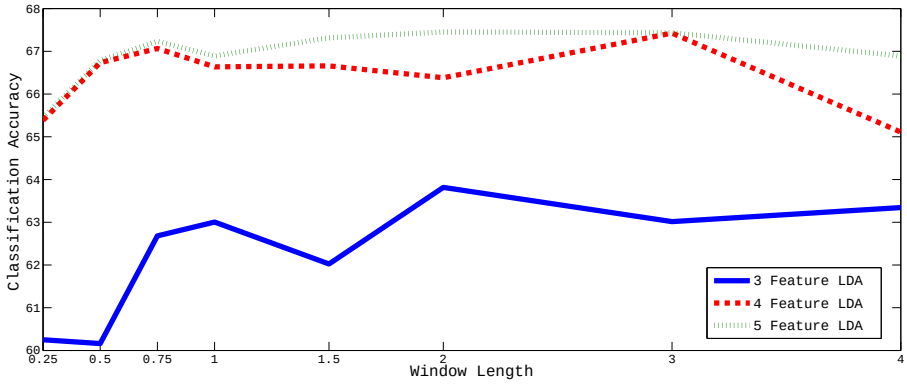
Ideally, classification of gyroscope signals to identify periods of artefact contamination in the EEG would permit participant independent head-movement artefact detection. This



(a) Sensitivity



(b) Specificity



(c) Accuracy

Figure 4.10: LDC results for each of the examined window lengths.

would allow an artefact detection system to be applied to an unseen participant, where the classifier has no prior knowledge of the user's EEG or movement characteristics as measured by the gyroscopes, and without the necessity of a calibration stage. The LDCs described in 4.4.2 were participant dependent, and although they verified the feasibility of using gyroscopes to detect EEG artefacts produced by head-movements, improved classification performance is sought for a practical artefact detection system. This is especially true considering that due to the high intra-person variability in EEG characteristics, participant independent systems typically perform considerably worse than their participant-specific equivalent. To achieve participant independent performance, support vector machine classifiers using the full feature set were investigated for this classification task. The justification for applying such an approach is centred on the fact that support vector machines classification have been shown to outperform linear discriminant classifiers for a range of classification tasks as demonstrated by Meyer et al. (2003); Lotte et al. (2007) and further demonstrated for the EEG artefact detection task in Chapter 3.

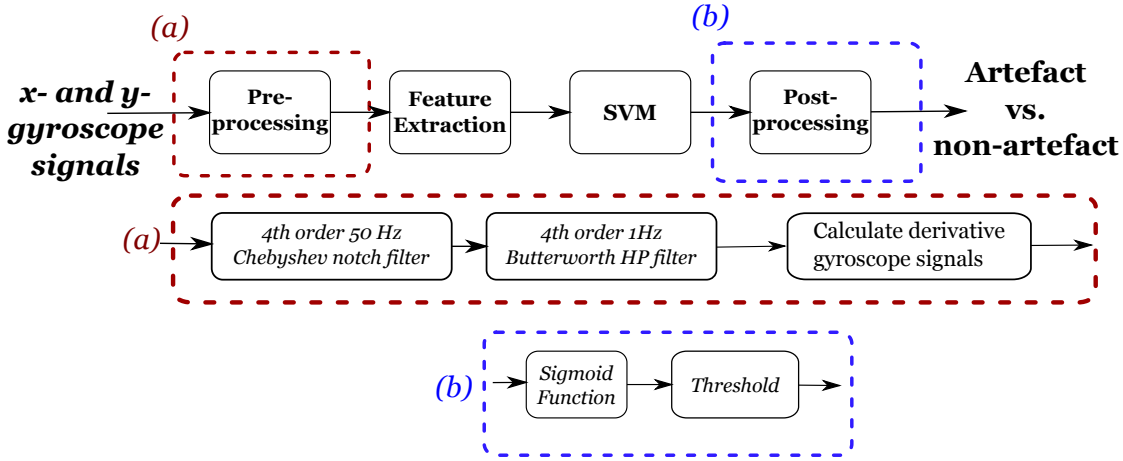


Figure 4.11: System architecture for gyroscope support vector machine head-movement artefact detector.

4.5.1 Gyroscope SVM

Support vector machines are a well-established pattern recognition technique for supervised learning, particularly suited to two-class classification problems (Cristianini and Shawe-Taylor, 2000). SVMs have been utilised for a wide range of classification tasks in the biomedical signal processing field in general and within the domain of automated EEG event detection in particular. A brief theoretical introduction to support vector machines is outlined in Section 3.5. In performing artefact detection using gyroscope sig-

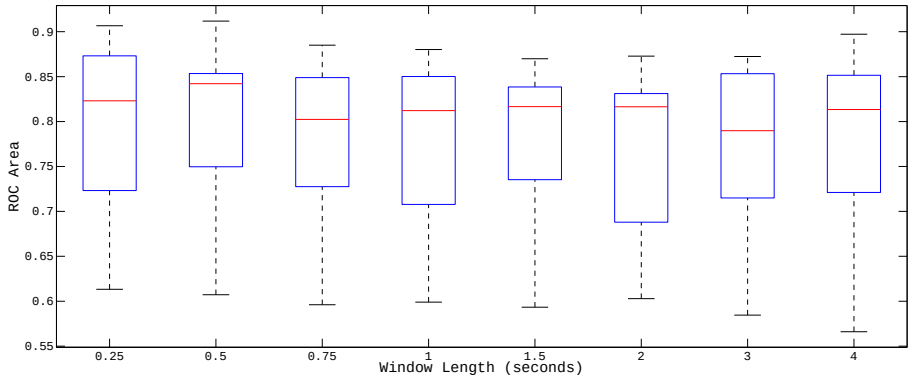


Figure 4.12: Boxplots of SVM ROC areas for each of the examined window lengths for the participant independent experiment.

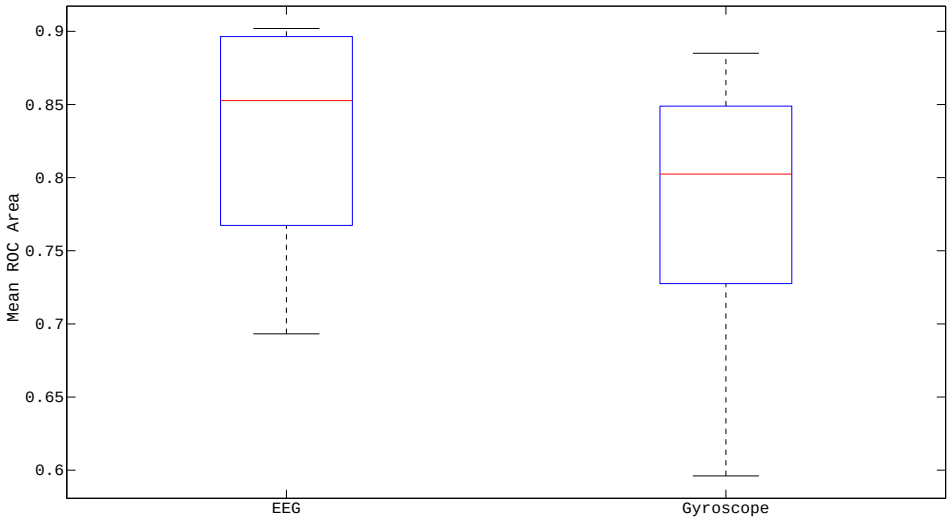
nals, Gaussian RBF kernel SVMs were used to perform the classification task. The SVM classification system is illustrated in Figure 4.11. The SVM classifiers were trained and tested using LOO-CV as described in Section 3.5. Participant independent classification was investigated for the aforementioned range of epoch lengths.

Boxplots of ROC areas across participants for the participant independent experiment for each window length investigated are displayed in Figure 4.12. Median ROC areas were typically between 80% and 85%, although there was considerable variability across participants. Despite this variability, the SVM classifiers guarantee ROC areas between 60 % and 90% for detecting artefacts using gyroscope signals in an unseen participant.

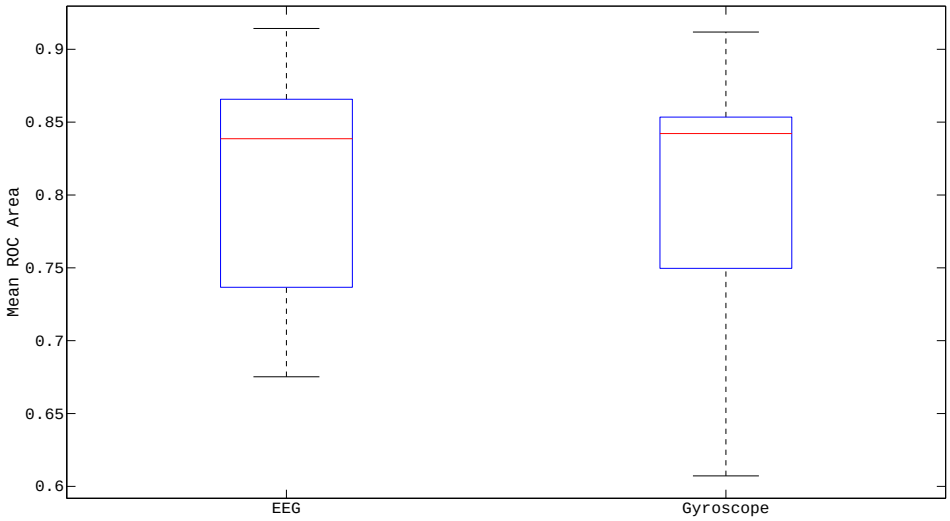
4.5.2 Comparison of EEG and gyroscope SVM performance

The feasibility of using gyroscope signals to predict if artefacts are produced in the EEG was confirmed with the use of SVM classifiers. Let us now consider the question: how does the gyroscope SVM compare to the EEG classifier from Chapter 3, in determining whether artefacts arising from head-movements have corrupted the EEG?

The performance of the EEG and gyroscope classifiers are compared in Figures 4.13a and 4.13b which show boxplots of mean ROC area across participants for window lengths of 0.5 (the best-performing gyroscope classifier) and 0.75 seconds (the best-performing EEG classifier), respectively. Performance of EEG and gyroscope classifiers were similar for both window lengths. For the 0.5 second window, the EEG classifiers had a higher median ROC area in addition to lower variability across participants. For the 0.75 second window, the



(a) 0.5 seconds was the top-performing window length for the gyroscope classification task.



(b) 0.75 seconds was the top-performing window length for the EEG classification task.

Figure 4.13: Boxplot of mean ROC areas for a window lengths of 0.5 and 0.75 seconds for the participant independent experiments. EEG and gyroscope results are shown.

median ROC area was higher for the gyroscope classifier, although the difference was slight.

4.6 Summary

In this chapter the feasibility of using features extracted from gyroscopes to detect head-movement artefacts in the EEG was investigated. Results from the mutual information calculations and linear discriminant classifiers indicated that the gyroscope feature set was useful in discriminating between resultant EEG artefact and non-artefact classes. LDC classification accuracies of between 65 % and 68 % for the 4- and 5- feature LDCs validated the use of gyroscope features to detect non-cerebral, movement-induced activity in the EEG. The classification results for the support vector machine classifiers showed good separation between gyroscope signals during normal EEG and those during head-movement artefact, providing a strong argument for including gyroscopes in an EEG artefact detection system. Indeed, the gyroscope classifiers performed at similar levels to the EEG classifiers (mean ROC areas between 80 % and 84 %) and even outperformed them for some window lengths. Once more, however, a single best window length was not chosen, due to the variability in mean ROC areas across participants for each window length investigated. The primary findings of this chapter can thus be summarised as follows: the production of artefact in the EEG can be predicted using only features extracted from the gyroscope signal and a suitably trained SVM classifier, and the accuracy of this prediction approaches that of an SVM classifier trained on EEG features. In the next chapter, information from both EEG and gyroscope signals will be combined to investigate if these different modalities offer complimentary information regarding the detection of head-movement artefacts in the EEG, and whether these physiological signals can be combined to improve artefact detection.

Chapter 5

Multimodal detection of EEG head-movement artefacts

5.1 Introduction

In the previous chapter, it was shown that statistical pattern recognition classifiers implemented with features extracted from gyroscope signals were effective in determining if head-movement artefacts were produced in the EEG. This poses the question: is complimentary information carried by the different modalities, and if so, how can this fact be exploited to improve the classification performance of the head-movement artefact detection task? In this chapter, a comprehensive data fusion analysis is conducted to investigate whether these different modalities carry complementary information, and if so, whether they can be combined to provide a more accurate detection of head-movement artefacts. To this end, several methods of combining these physiological and physical signals at the feature, decision and fusion levels are examined for their effectiveness in detecting EEG artefacts arising from head-movements.

Two journal papers (O'Regan et al., 2013a; O'Regan and Marnane, 2013) have arisen from the work outlined in this chapter.

5.2 Multimodal fusion

Fusion of information from EEG and gyroscope signals can be performed at the feature level or at the classifier level. Feature fusion (sometimes referred to as early integration) can take various forms, although typically the most straightforward feature fusion method is utilised and features from different modalities are concatenated on a per-epoch basis. Fusion at the classifier level (sometimes referred to as late integration) employs separate classifiers for each signal and combines the results thereafter. Classifier fusion is a well-researched method of combining information from different modalities, with a large range of options available to combine the outputs from a set of classifiers into a final binary decision (Kittler et al., 1998; Kuncheva, 2004; Mandic et al., 2005). There are many examples in the literature where classification using a combination of EEG and additional physiological signals outperform the individual, base classifiers in the experiment (Peng et al., 2007; Kapoor et al., 2008; Qian et al., 2009; Polikar et al., 2010). However, the results reported in the literature for combining EEG and additional physiological signals conflict in their recommendations for early or late integration of signals. Accordingly, they do not clearly indicate whether early or late integration will offer better performance for the multimodal head-movement artefact classification task. In the seizure detection domain, Greene et al. (2007) investigated the combination of EEG and ECG signals in improving the performance of neonatal seizure detection and found feature fusion to outperform classifier fusion. Malarvili and Mesbah (2008) found the opposite with fusion of EEG and ECG signals at the classifier level offering better detection of seizure in neonatal EEG. Bermudez et al. (2007) performed a similar analysis for detecting temporal lobe epilepsy in adults but found that classifier fusion offered better classification performance than feature fusion. The work in this chapter therefore investigates what method of multi-modal data fusion is most effective in combining EEG and gyroscope signals for the detection of head-movement artefacts.

In Chapters 3 and 4 it was discovered that choice of window length was largely unimportant for EEG and gyroscope classifier performance. With this in mind, coupled with the fact that the convention in the literature has been to utilise window lengths of one second, the multimodal fusion experiment outlined in this chapter was carried out for window lengths of one second.

5.3 Feature fusion

Perhaps the most straightforward method of fusing information from different modalities is that of feature fusion. In this process, features extracted from the individual signals are combined before being applied to a classifier. While there are several methods of performing this combination, the most common is to simply concatenate features extracted from the different modalities. In support vector machines classification, the SVM is characterised by the normal to the maximum-margin hyperplane, where its components specify the weights (i.e. importance) of the features, such that higher absolute values have a larger impact on the decision function. Thus SVMs assign weights to each feature all on its own. In this context, unweighted feature concatenation is an appropriate feature fusion technique to combine the EEG and gyroscope signals for the head-movement artefact classification task. Thus, for a given epoch of an EEG channel, a feature set of 69 features is extracted from the EEG, and a further 80 features are obtained from the corresponding epoch of the gyroscope signals. A feature set of 149 features is then associated with each epoch of each EEG channel. By this means, gyroscope information is incorporated into per-channel EEG classification.

The EEG and gyroscope features were used to train a radial basis function SVM to separate between sections of EEG that contained head-movement artefact and sections that were artefact-free. The training and testing process was the same as those outlined for the individual EEG and gyroscope classifiers in Chapters 3 and 4 and the reader is referred there for more details. Once more, to achieve participant independent artefact detection, LOO-CV was used. An illustration of the feature fusion architecture is outlined in Figure 5.1.

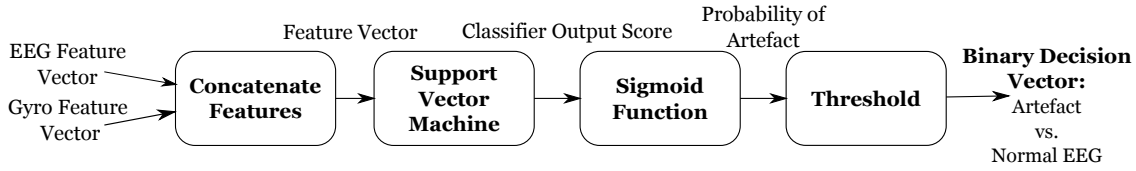


Figure 5.1: Overview of the feature fusion classifier combination architecture. Feature fusion as illustrated in the diagram is repeated for each EEG channel.

5.3.1 Feature fusion performance

Figure 5.2 displays the boxplots of the ROC areas across participants for the EEG, gyroscope and feature fusion classifiers. Both mean and median ROC areas of the feature fusion classifier are higher than those of the EEG and gyroscope classifiers. This indicates that the EEG and gyroscope features contain complementary information in discriminating between head-movement artefact and background EEG. Another trend that is visible in Figure 5.2 is that the variability of ROC areas across participants is smaller for the feature fusion classifier than those of the individual EEG and gyroscope SVMs. This indicates that complementary information from the EEG and gyroscope signals offers more robust classification performance across participants than using either of the individual EEG or gyroscope classifiers in isolation.

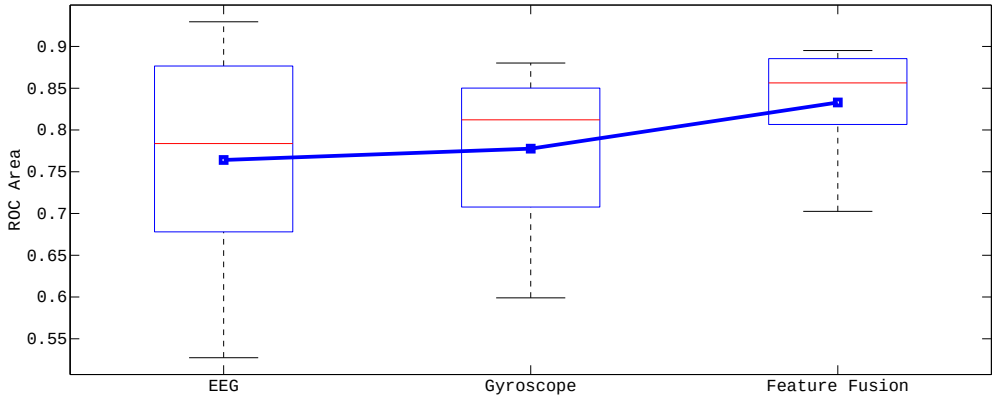


Figure 5.2: Boxplot of ROC areas for feature fusion classifier with an epoch length of one second. Boxplots of the EEG and gyroscope classifiers are also displayed. Mean ROC areas are shown by the blue line.

5.4 Score fusion

In the score fusion method of classifier combination, EEG and gyroscope classifiers are trained separately and combined afterwards. Fusion takes the form of combining the gyroscope classifier output probabilities with the output probabilities from each EEG channel to allow per-channel artefact detection (Figure 5.3). In this thesis, score fusion methods using non-trainable, fixed combining rules are investigated. These fixed rule combiners are available for combination as soon as the base classifiers are trained and

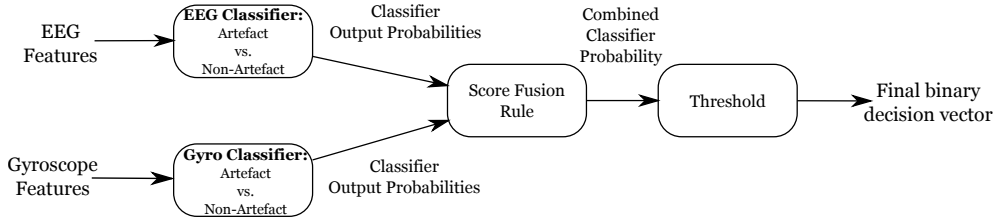


Figure 5.3: Overview of the score fusion classifier combination architecture. The process illustrated in the diagram is repeated for each EEG channel.

make use of the fact that the outputs of the EEG and gyroscope classifiers have a clear interpretation, namely, they are the class posterior probabilities. The combined classifier $Q(\mathbf{x})$ is then:

$$Q(\mathbf{x}) = \mathcal{F}[C(\mathbf{x})] \tag{5.1}$$

where $Q(\mathbf{x}) = Q_j(\mathbf{x});\ j = 1, ..., c$ for j base classifiers C, and $\mathcal{F}$ is a fixed-rule combination function. Several common, fixed-rule combination functions are investigated in this work: mean, product, maximum and minimum rules. Kittler et al. (1998) and Tax et al. (2000) showed that there are advantages and disadvantages to score fusion carried out with each of these rules. This study examines how these characteristics unfold in the context of EEG head-movement artefact detection.

5.4.1 Mean/Sum rule

The mean rule is simply the average of the individual, base classifier posterior probabilities:

$$Q(\mathbf{x}) = \frac{1}{L}\sum_{j}^{L} C_j(\mathbf{x}) \tag{5.2}$$

for j classifiers. In combining the EEG and gyroscope classifiers, we thus get:

$$P_{Artefact}(\mathbf{x}) = \frac{P_{EEG}(\mathbf{x}) + P_{Gyro}(\mathbf{x})}{2} \tag{5.3}$$

where $P_{EEG}(\mathbf{x})$ is the per-channel probability of head-movement artefact existing on a given EEG channel, and $P_{Gyro}(\mathbf{x})$ is the probability of head-movement artefact existing on the same EEG channel, as estimated by the gyroscope classifier. In some papers, the scaling factor L is omitted, and the combination output $Q(\mathbf{x})$ becomes the sum of the posterior probabilities (Kittler et al., 1998; Kittler and Alkoot, 2001).

5.4.2 Product rule

Combined classifier posterior probabilities are created via the product rule by multiplying the output probabilities from each of the individual classifiers, such that:

$$Q(\mathbf{x}) = \prod_j C_j(\mathbf{x}) \tag{5.4}$$

The product rule has been shown to produce good results if the individual classifiers are independent, e.g. if the classifiers have been created for different feature spaces that are entirely unrelated. The product rule assumes that the confidence estimates are reliable and noise-free; this is assumed to be the case for the EEG and gyroscope SVM classifiers due to the use of LOO-CV in training and testing. It fails if these confidence estimates are very small or zero. The product rule for combining the EEG and gyroscope classifiers is thus given by:

$$P_{Artefact}(\mathbf{x}) = P_{EEG}(\mathbf{x}) \times P_{Gyro}(\mathbf{x}) \tag{5.5}$$

5.4.3 Maximum rule

The maximum rule selects the classifier that is most confident of the classification decision, i.e. the combined classifier posterior probability $Q(\mathbf{x})$ is that of the individual classifier with the highest output probability.

$$Q(\mathbf{x}) = \max_j \{C_j(\mathbf{x})\} \tag{5.6}$$

While this rule may be perceived as the most intuitive (choosing the most expert opinion for a given decision), it is subject to a number of potential pitfalls. Most notably, the maximum rule can fail if one of the classifiers is overtrained; the overtrained classifier is then overconfident of its decision, dominating the outcome, without improving the classification performance. The use of SVMs with parameters and model selection performed with LOO-CV for both EEG and gyroscope classifiers should mitigate this possibility. The maximum rule also underperforms if there are simple component classifiers that are insensitive to nuances in the classification task that more complicated classifiers are able to detect. The insensitive classifiers then tend to dominate the maximum rule, reducing performance. The maximum rule for combining the EEG and gyro classifiers is:

$$P_{Artefact}(\mathbf{x}) = \max[P_{EEG}(\mathbf{x}), P_{Gyro}(\mathbf{x})] \tag{5.7}$$

5.4.4 Minimum rule

The minimum rule selects the outcome of the classifier that has the least objection against a certain class. It is a conservative combination rule, with the posterior probability $Q(\mathbf{x})$ that of the individual classifiers with the lowest output probability:

$$Q(\mathbf{x}) = \min_j \{C_j(\mathbf{x})\} \tag{5.8}$$

This is displayed for combining the EEG and gyroscope classifiers in Equation 5.9:

$$P_{Artefact}(\mathbf{x}) = \min[P_{EEG}(\mathbf{x}), P_{Gyro}(\mathbf{x})] \tag{5.9}$$

In the minimum rule, it is the least confident classifier that dictates the combined posterior probability; a measure which takes precautions against overtraining in one (or more) of the individual classifiers. However, for this very reason, the minimum rule can be undermined by one uncertain, or poorly discriminative component classifier. As was seen in Chapters 3 and 4, both EEG and gyroscope SVM classifiers were useful in discriminating between artefact and non-artefact in the EEG signal.

5.4.5 Score fusion performance

Figure 5.4 illustrates that combining the individual EEG and gyroscope classifiers using any of the fixed-rule score fusion combinations results in improved performance in the head-movement artefact detection task. For each of the score fusion combination rules, the mean ROC area is higher than those of either the EEG or gyroscope classifiers alone. Similarly, the median ROC areas for each of the score fusion classifiers is higher than that of the EEG classifier and equal to or greater than that of the gyroscope classifier. As was the case for the feature fusion classifier, each of the score fusion combination rules, provides a lower variability in ROC areas that adds a level of robustness to the head-movement classification task. Amongst the score fusion combination rules, the mean/sum rule provides the greatest improvement in classifier performance. This is further highlighted in Figure 5.5, which displays the mean ROC plots for each of the investigated score fusion combination rules as well as those of the EEG and gyroscope classifiers, where the sum rule can be seen to dominate (i.e. lie above and to the right of) each of the other classifiers. It is also evident that the performance of the other combination rules match that of the sum rule over defined sections of the ROC plot. For example, in applications where head-movements

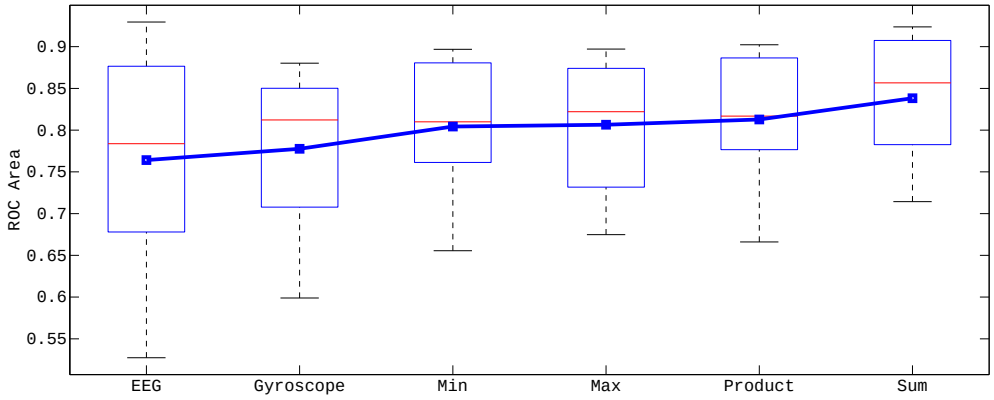


Figure 5.4: Boxplot of ROC areas for score fusion classifiers with an epoch length of one second. Boxplots of the EEG and gyroscope classifiers are also displayed. Median ROC areas are shown within the boxplots by the red lines. Mean ROC areas are shown by the blue line.

must be detected with a high degree of accuracy (i.e. high specificity), deploying the product or minimum rule in place of the sum rule will suffice. Similarly, for applications where detecting as many artefacts as possible is of concern (i.e. high sensitivity), utilising the maximum rule will provide the same performance as that of the sum rule. However, as there exists little or no difference in computational complexity between the rules, if score-level fusion of classifiers is to be employed then the sum rule should be used. This result is consistent with findings from other classifier combination tasks. Kittler et al. (1998) found that for an identity verification task, that amongst score fusion combination rules, the sum rule is most resilient to estimation errors.Tax et al. (2000) showed that in a two-class problem in which the posterior probabilities are well estimated and with completely independent feature spaces, the product rule performs best. However, when these assumptions are violated, the sum rule outperforms the product rule. In the head-movement artefact classification task, it may be unreasonable to assume that the EEG and gyroscope signals are independent. Although the feature spaces are derived from different physiological signals, the class labels for both modalities are dependent upon the EEG signal.

5.5 Decision fusion

As was the case for the score fusion methods, the EEG and gyroscope classifiers were trained and tested separately and combined on a per-channel basis afterwards. Decision

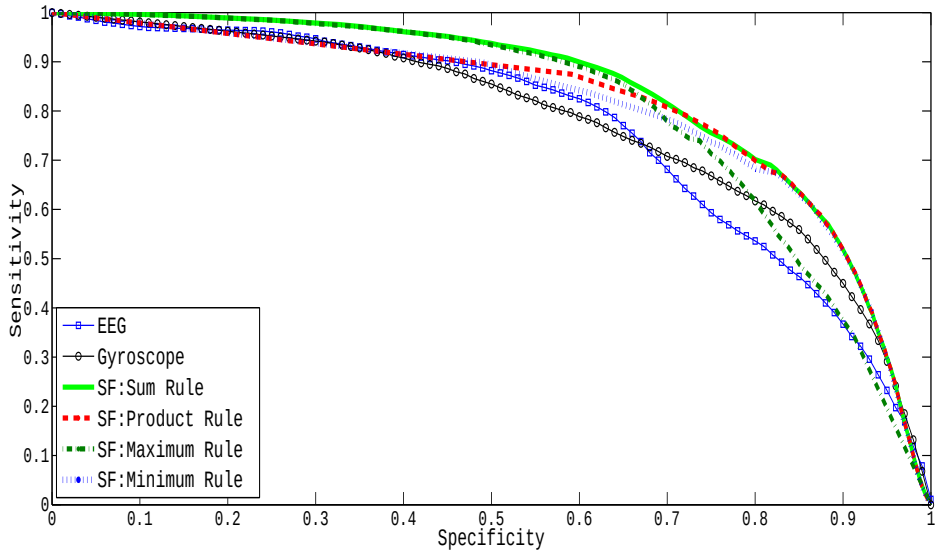


Figure 5.5: Mean ROC plots for score fusion classifier combinations. Mean ROC plots for EEG and gyroscope classifiers are also displayed.

level classifier fusion is accomplished by combining the binary output decisions $\mathcal{D}_i(\mathbf{x})$ from the individual classifiers (after the threshold has been applied to the classifier output probabilities), where the binary decisions are represented by Equations 5.10 and 5.11. The decision fusion classification architecture is illustrated in Figure 5.6. In the decision fusion methods investigated in this thesis, the same threshold was applied to the EEG classifier and to the gyroscope classifier, such that $\theta_{EEG} = \theta_{Gyro}$. While an optimal set thresholds $[\theta_{EEG}, \theta_{Gyro}]$, the thresholds were varied together so as to reduce the number of parameters in the system. Two decision fusion methods are investigated: logical OR and logical AND.

$$\mathcal{D}_{EEG}(\mathbf{x}) = \begin{cases} 0 & \text{EEG classifier decides normal EEG} \\ 1 & \text{EEG classifier decides artefact} \end{cases} \tag{5.10}$$

$$\mathcal{D}_{Gyro}(\mathbf{x}) = \begin{cases} 0 & \text{Gyroscope classifier decides normal EEG} \\ 1 & \text{Gyroscope classifier decides artefact} \end{cases} \tag{5.11}$$

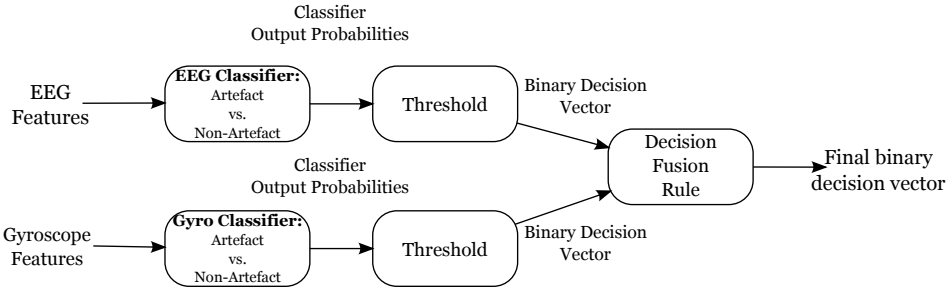


Figure 5.6: Overview of the decision fusion classifier combination architecture. The decision fusion process highlighted in the diagram is repeated for each channel of EEG.

5.5.1 Logical AND method

This classifier fusion method is performed using the logical AND operation, with 1 indicating the presence of artefact. The combined classifier output $\mathbf{D}(\mathbf{x})$ is then:

$$\mathbf{D}(\mathbf{x}) = \mathcal{D}_{EEG}(\mathbf{x}) \;\; \mathbf{AND} \;\; \mathcal{D}_{Gyro}(\mathbf{x}) \tag{5.12}$$

If 1 exists in both classifier decisions, then the result of the combination of the two decision vectors will also be a 1, indicating that artefact is present in the epoch in question. Thus, if an artefact is to be detected using the logical AND method, both classifiers must detect the presence of the artefact individually. In this light, the logical AND classifier combination can be thought of as producing classification decisions with high specificities.

5.5.2 Logical OR method

This decision fusion technique examines the combination of EEG and gyroscope classifier binary output decisions using the logical OR operator. The combined classifier output $\mathbf{D}(\mathbf{x})$ is:

$$\mathbf{D}(\mathbf{x}) = \mathcal{D}_{EEG}(\mathbf{x}) \;\; \mathbf{OR} \;\; \mathcal{D}_{Gyro}(\mathbf{x}) \tag{5.13}$$

If the binary decision vector from either EEG or gyroscope classifier detects artefact, the result of the logical OR classifier combination will report artefact. Consequently, the logical OR decision fusion rule should produce classifier decisions with high sensitivities.

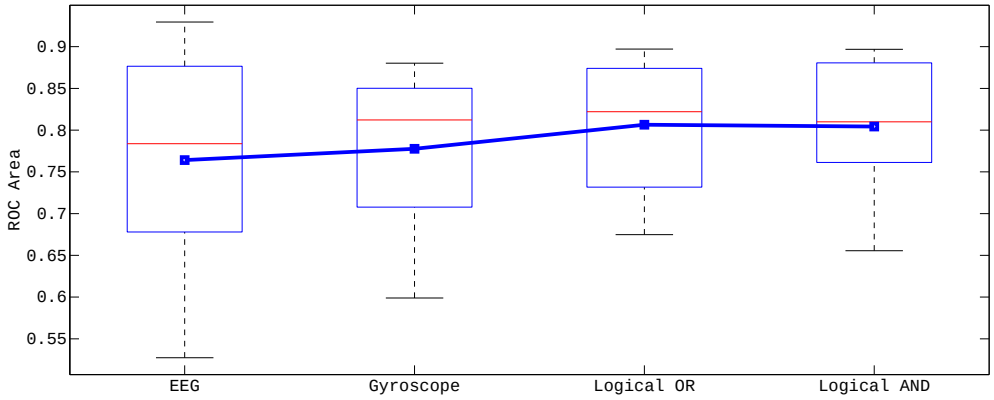


Figure 5.7: Boxplot of ROC areas for decision fusion classifiers with an epoch length of one second. Boxplots of the EEG and gyroscope classifiers are also displayed. Mean ROC areas are shown by the blue lines.

5.5.3 Decision fusion performance

Boxplots of the ROC areas across participants for the decision fusion classifiers are presented in Figure 5.7. Both logical OR and logical AND fusion methods outperform the EEG classifier as measured by mean and median ROC areas. Mean ROC plots for the logical OR decision fusion method and the logical AND combination are displayed in Figure 5.8. As expected, the logical OR combination rule is highly sensitive to the detection of artefacts; if either EEG or gyroscope classifier detects a head-movement artefact, the logical OR output will flag that EEG epoch as artefact. The logical OR combination rule thus outperforms the logical AND classifier and the individual EEG and gyroscope classifiers for specificities below 0.65. However, this increased sensitivity comes at a significant reduction in performance for high specificities, where the logical OR rule falls below those of the individual EEG and gyroscope classifiers. In contrast to this, the logical AND combination rule is highly conservative and only marks a section of EEG as artefact if both EEG and gyroscope classifiers agree, and consequently performs best if high specificities are required. In this regard, the logical OR and the logical AND are analogous to the score fusion maximum and minimum rules, respectively.

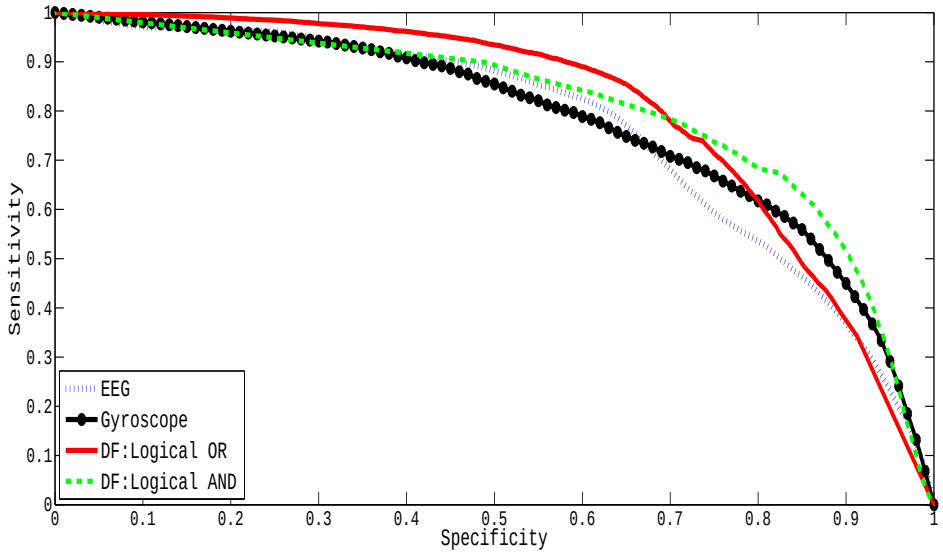


Figure 5.8: Mean ROC plots for decision fusion classifier combinations. Mean ROC plots for EEG and gyroscope classifiers are also displayed.

5.6 Comparison of fusion methods: early integration vs. late integration

The feature fusion and top-performing score and decision fusion classifiers are compared in Figure 5.9. As the decision-level fusion offered similar classification performance to the minimum and maximum score fusion combination rules, it is unsurprising that the sum rule outperforms the best decision fusion classifier. The mean and median ROC areas of the feature fusion and sum rule score fusion are approximately equal. However, the variability in ROC areas of the sum rule is larger than that of the feature fusion. Accordingly, for the head-movement classification task where EEG and gyroscope features are available, combination of these signals at the feature level is advised. However, for classifiers other than support vector machines, score fusion may be more appropriate. SVMs do not suffer from the "curse of dimensionality" to the same extent as many other classifier models; hence, combining the different modalities at the feature level is unlikely to lead to excessive overfitting. For classifiers that are highly susceptible to overfitting, such as neural networks, feature fusion may do just this. It is proposed that this may indeed be the case in the systems outlined by Malarvili and Mesbah (2008) and Bermudez et al. (2007), where neural networks were deployed and classifier fusion was advocated in place of feature fusion.

In practice there may be inherent advantages or disadvantages to fusing data from different modalities by means of feature or score fusion techniques. There exists a number of scenarios whereby use of the score fusion sum rule makes more sense. Score fusion offers the advantage that the artefact detection system will continue to operate if the gyroscope signal becomes unavailable. In the experience of the Biomedical Signal Processing Group at U.C.C. in collecting clinical data, it is often the case that additional physiological signals are not recorded or partially recorded, particularly during longer recordings. In such an eventuality, the classifier combination stage outlined in this paper can be switched off, and the EEG artefact detection classifier can continue alone. Were the gyroscope to become unavailable in the feature fusion case, the system would not be able to operate, as the gyroscope features would now lie in an unexpected region of the feature space, leading to likely adverse classification performance. Of course this feature requires the recognition by the clinician/researcher that the gyroscope signal is unavailable or corrupted. This step would be automated if the maximum score fusion or the logical OR decision fusion rules were adopted, albeit at a loss in classification accuracy compared to feature fusion or score fusion, sum rule. Furthermore, in adapting the score fusion method, additional modalities can be easily incorporated into the classification task without re-training of the classifiers. Thus, if additional physical signals such as accelerometer data or electromyogram signals were to be made available, score fusion combination could easily accommodate these. For marginally better classifier performance, feature fusion would require the classifier to be entirely re-trained incorporating features from the added physiological signals.

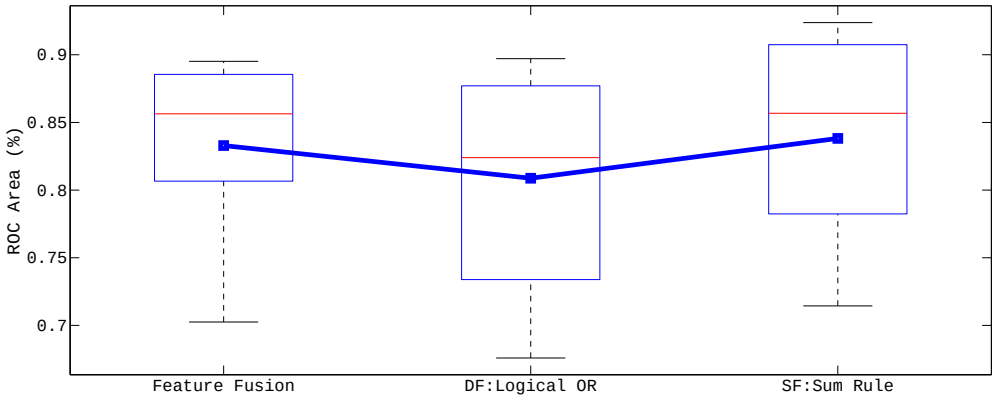


Figure 5.9: Boxplot of ROC areas for the best performing feature, score and decision fusion classifiers. Boxplots of the EEG and gyroscope classifiers are also displayed. Mean ROC areas are shown by the blue line.

5.7 Summary

In light of examining a range of multimodal classifiers incorporating both EEG and gyroscope signals, it was found that the fusion of signals at both feature and classifier levels improved detection of head-movement artefact in ambulatory EEG when compared to using either EEG or gyroscopes alone. This result was observed for feature fusion as well as for each of the score-level and decision-level fixed-rule combinations that were investigated. Amongst these methods, it was found that feature fusion and the score-level, sum rule offered the best classification performance by increasing the ROC areas and reducing inter-participant variability. Improvements in mean ROC area for a window length of one second was 8 % for the feature fusion and score fusion sum-rule combination methods when compared to the EEG classifier alone. These results confirm the complimentary nature of information carried by these different modalities. Thus, in order to most effectively detect head-movement artefacts in EEG, a classifier that combines EEG and gyroscope features at either the feature or score level is recommended.

Chapter 6

Robust epileptiform activity detection

6.1 Introduction

Automatic detection of epileptiform activity (spikes, sharp waves and spike-wave complexes) in the electroencephalogram (EEG) to speed up and disambiguate EEG analysis in the diagnosis of epilepsy has been a highly desired research goal for over 40 years. However, a major obstacle in the deployment of automated epileptiform activity detection systems stems from contamination of the EEG by electrical activity arising from non-cerebral sources. In automatic epileptiform activity and epileptic seizure detection, these artefacts can be misinterpreted as epileptiform activity, leading to an unacceptably large number of false positive detections. In the previous chapters, several artefact detection algorithms were investigated for their ability to detect the presence of head-movement artefact in ambulatory EEG. In this chapter, information from artefact detection classifiers similar to those in the previous chapter, is combined with neurological event detection classifiers to improve the performance of the neurological event detection system. Accordingly, a novel artefact processing system is outlined, whereby the detection of ocular and movement artefacts is performed in parallel to epileptiform activity detection. Fusion of these support vector machines classifiers is investigated with results showing a considerable reduction in the number of false detections whilst continuing to accurately detect epileptic seizure and short-duration, interictal epileptiform events. Two journal submissions (O'Regan et al., 2013b; Kelleher et al., 2013) and an international peer-reviewed

conference publication (Kelleher et al., 2011) have arisen from the work outlined in this chapter.

6.2 Automated epilepsy detection

Due to the sub-clinical nature of many epileptic seizures, the electroencephalogram (EEG) is the primary tool used in the diagnosis of epilepsy. Diagnosis is made based on the presence of epileptic seizure activity (Figure **??**), which can last from several seconds to an hour, and also based on the presence of shorter duration interictal, epileptiform activity (Figure **??**), which is an indicator of susceptibility to epileptic seizure (Chatrian et al., 1964).

Following a referral from a neurologist, the initial test for epilepsy is known as a routine EEG, which lasts 20 - 40 minutes and looks for signs of abnormal waveforms in the recording. These abnormal waveforms comprise some or all of the following features: slow waves, spikes and sharp waves, spike-wave complexes, and polyspikes. If, at the end of the test, no such activity has been recorded and epilepsy is still suspected, the patient may be asked to return to the hospital for a longer, continuous EEG recording (usually lasting 24 - 72 hours). The recorded signal is then visually inspected by a neurophysiologist for abnormalities. Due to the costly, time-consuming nature of continuous EEG monitoring, it has been proposed that an ambulatory EEG system, with the capability to automatically detect epileptic seizure activity as well as the shorter duration epileptiform activity, would be quite useful for clinical and domestic monitoring of patients (Waterhouse, 2003; Gotman and Gloor, 1976; Indiradevi et al., 2008). Such a system would help the clinician pinpoint the exact location of abnormalities in the recording, thus vastly reducing post-recording analysis time.

6.3 Ocular and movement artefacts

Contamination of the EEG by electrical activity arising from non-cerebral sources is a major obstacle in the achievement of the dual criteria of automatically detecting almost all epileptiform events while maintaining a low false detection rate. In particular, head-movement and ocular artefacts (Figures 6.4 and 6.5) can obscure, and in the case of ocular artefacts, mimic the short duration epileptiform activity. In automatic epileptiform activ-

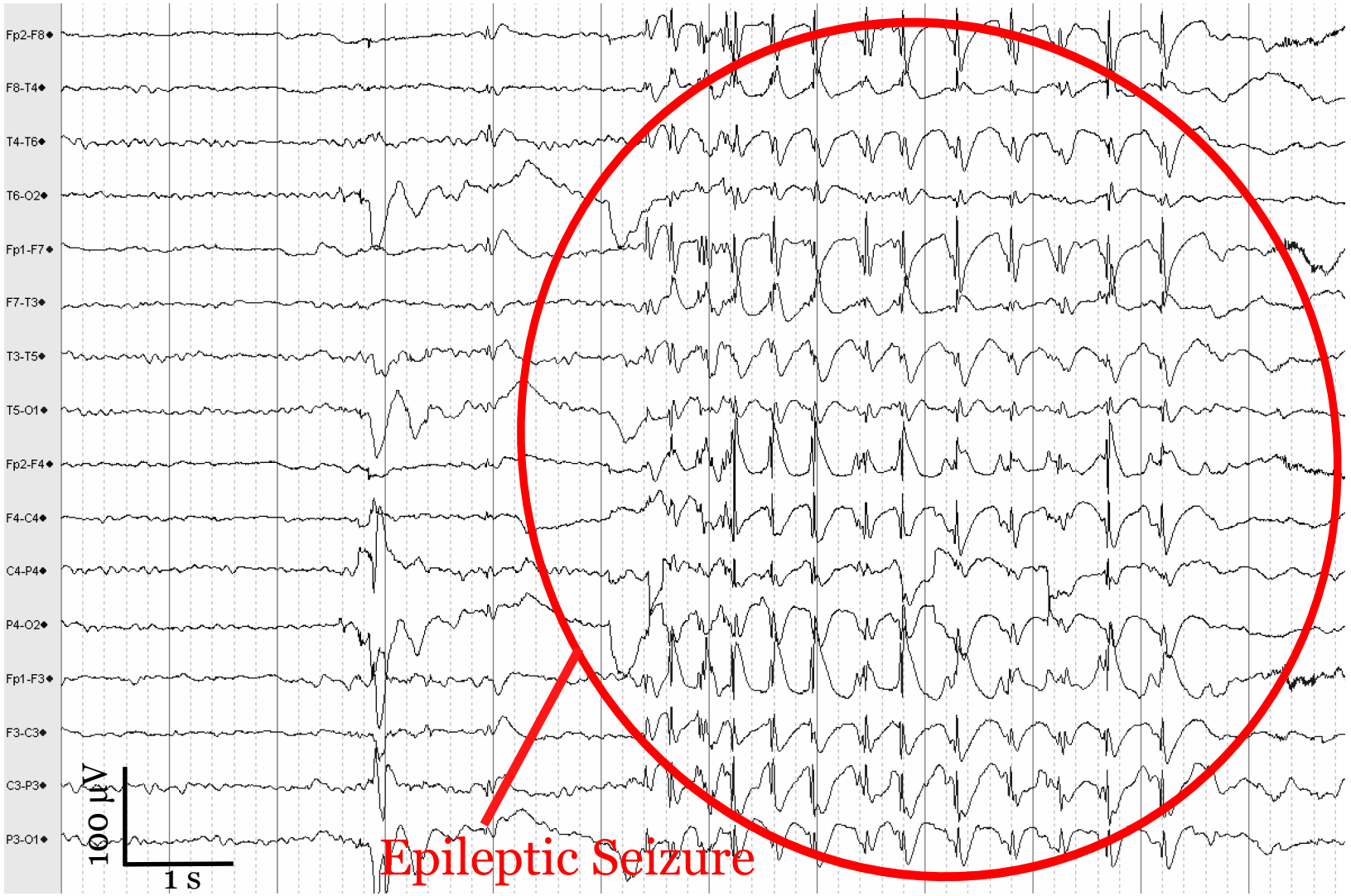


Figure 6.1: EEG showing an example of a generalised idiopathic epileptic seizure recorded from an adult patient at CUH. The data was recorded using a Viasys Nicolet EEG machine with a sampling frequency of 256 Hz.

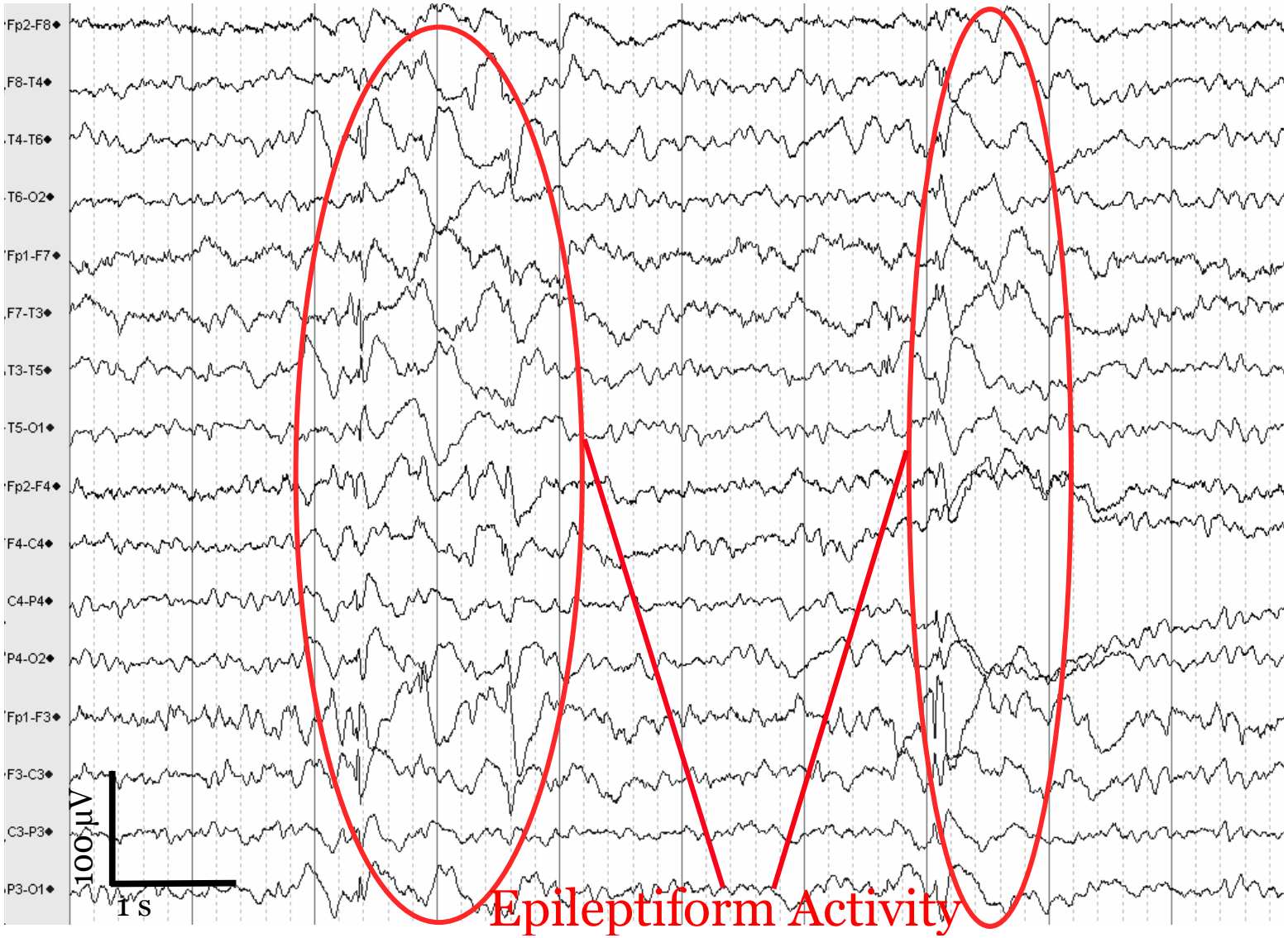


Figure 6.2: EEG sample showing two instances of epileptiform activity recorded from an adolescent patient at CUH. The data was recorded using a Viasys Nicolet EEG machine with a sampling frequency of 256 Hz.

ity and epileptic seizure detection systems, these artefacts are frequently misinterpreted as epileptiform activity, leading to an unacceptably large number of false positive detections (Figure 6.3).

Ocular artefacts are the most prevalent artefact found in many EEG recordings and are a major contributing factor to the inaccurate classification of epileptiform activity (Kelleher et al., 2010). The amplitude of ocular artefacts can be several times larger than brain scalp potentials, thus seriously interfering with the EEG recording. As the eyeball moves, the potential difference that exists between the cornea and retina changes, producing the electrooculogram (EOG) signal that frequently appears on the EEG as ocular artefact. This ocular artefact signal propagates across the scalp, rapidly diminishing with distance travelled from the eyes and is therefore most acute on the frontal electrodes (Lins et al., 1993; Picton et al., 2000). A number of examples of eye blinks and lateral eye movements are displayed in Figure 6.5.

As discussed in Chapter 2, in many clinical EEG trials, contamination by artefacts is minimized by controlling the test situation to limit movement. In an ambulatory setting restricting movement is both unrealistic, and in cases such as diagnosing epilepsy, may even be undesirable. However, these head movements can introduce a wide range of non-cerebral electrical activity into the EEG; typically, contamination is in the form of some combination of muscle (EMG), electrode pop and electrode movement artefacts (Figure 6.4). As seen in the previous chapters, muscle or EMG artefacts are predominantly high frequency signals whose amplitudes can vary depending on the muscle in questions as well as the strength of muscle firing (Goncharova et al., 2003). Electrode pop, which occurs with momentary loss of contact between the electrode and the scalp causes an abrupt impedance change, morphologically appearing as single or multiple sharp waveforms. These sharp waveforms are typically high amplitude vertical transients which are usually confined to a single electrode. Electrode movement occurs when the electrode moves with respect to the scalp. These movements can produce high-amplitude deflections in the EEG of the order of millivolts. Unlike electrode pop, EEG deflections caused by electrode movements are slower, often in the range of 1 - 10 Hz (van de Velde et al., 1999); these slower movements can then mimic epileptic seizure.

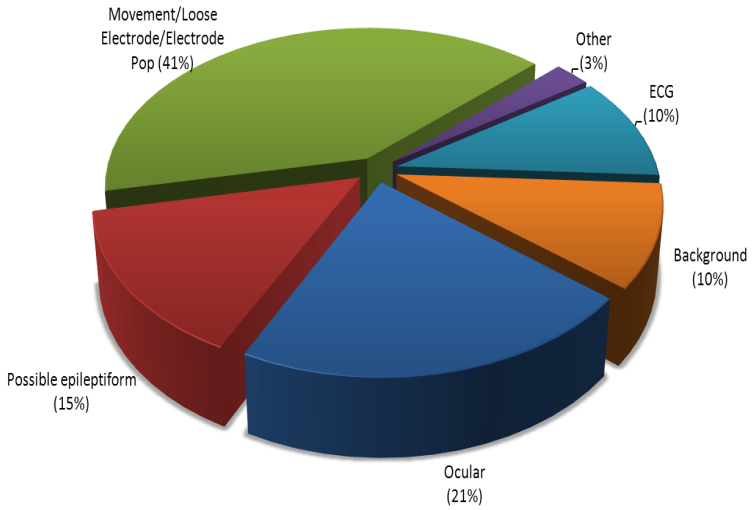


Figure 6.3: Breakdown of false detections by underlying cause in epileptiform activity detection classifier.

6.4 Reducing false positives without removing valuable EEG information

In the automated epileptiform activity detection system created by Kelleher et al. (2010), and indeed in the majority of systems reviewed in Section 2.8.1, it is envisaged that the EEG clinician will only review events detected by the algorithm; any false positives will be inspected and subsequently excluded by the clinician. Therefore, if the detection system fails to find an epileptiform event, then that section of EEG will not be presented to the clinician who will of course miss it. From a clinical perspective, this is unacceptable; the consensus amongst the clinicians involved in EEG data collection for this thesis is that they would be extremely reluctant to accept a tool that discards potentially useful epileptiform information. While a few undetected events may be tolerable, it is inappropriate to miss greater than $\sim 10\%$ of events, as clinicians will be reluctant to discard so much (important) information. This highlights one of the key sticking points in implementing statistical pattern recognition and machine learning algorithms in a medical environment. Given, the short duration of the epileptiform events in question, a GDR of 0.9 with FD/h in the low hundreds, could be considered impressive by machine learning standards, yet, is notably sub-standard for the desired medical application. However, this obstacle cannot be simply circumvented by increasing the GDR to 1 as this will lead to substantially more false detections. Consequently, if the system is frequently flagging non-epileptiform

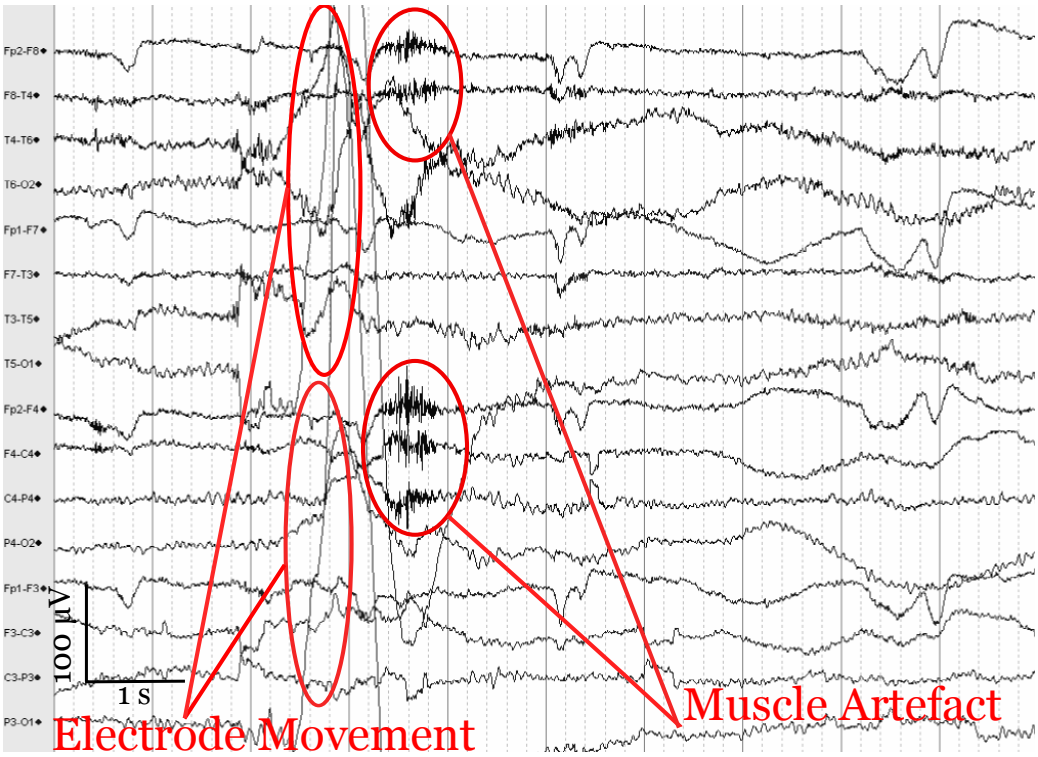


Figure 6.4: EEG sample from Patient 2 showing artefacts arising from head movements. The data was collected in the Department of Neurology at CUH using a Viasys Nicolet EEG machine using a sampling frequency of 256 Hz.

EEG as being epileptiform activity or epileptic seizure (i.e. a high false detection rate) then the clinician may turn off the automated EEG system and resume annotating the recording manually. It is therefore necessary that an automated epileptiform detection system detects all (or almost all) epileptiform events while keeping the number of false detections at a minimum.

By increasing the number of false positive detections, artefacts are major barrier to the deployment of automated epileptiform detection systems. The measures taken by several groups to reduce these false positives are described in Section 2.8.1. In summary, artefact rejection in automated epileptic seizure and epileptiform detection systems to date have focused primarily on rule-based thresholding methods. As SVMs and the fusion methods outlined in Chapters 3 and 5 have been shown to be highly successful for artefact detection, it is proposed that performing artefact detection using support vector machines classifiers on sections of EEG that contain suspected epileptiform activity may be an effective, alternative method of dealing with artefacts in automated epileptiform detection system. In this chapter a novel, artefact detection architecture is proposed, whereby only EEG that has been flagged as potentially containing epileptiform activity or epileptic seizure

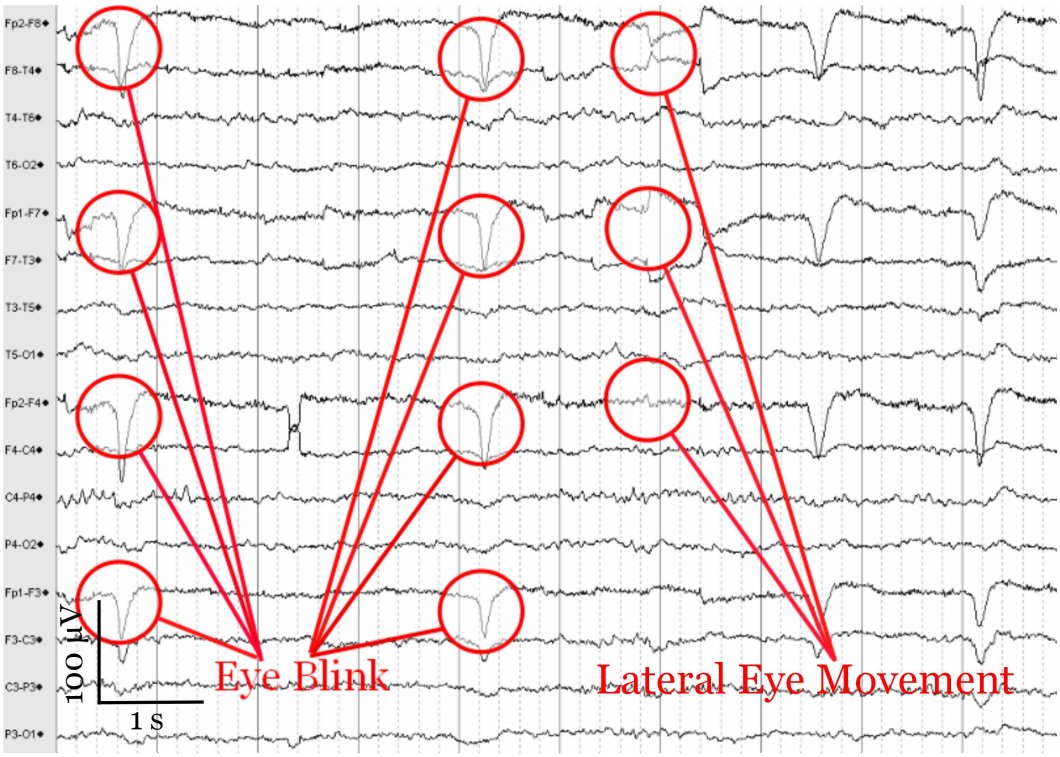


Figure 6.5: EEG sample from Patient 1 showing a number of instances of eye blinks and lateral eye movements. It is evident that the artefacts are most pronounced on frontal electrodes. The data was collected in the Department of Neurology at CUH using a Viasys Nicolet EEG machine using a sampling frequency of 256 Hz.

are subjected to an additional classification stage in order to reduce false detections due to artefacts.

6.5 Epileptiform detection classifier

The data used for this study consists of multi-channel EEG recordings obtained from 8 patients, each suffering from idiopathic generalised epilepsy and is outlined in detail in Section 2.9.2.

6.5.1 Detection system

The epileptiform activity and epileptic seizure detection system used in this study is a modification of the system used in the detection of neonatal seizures and has been described in detail by Kelleher et al. (2010, 2011); Temko et al. (2011b). An outline of

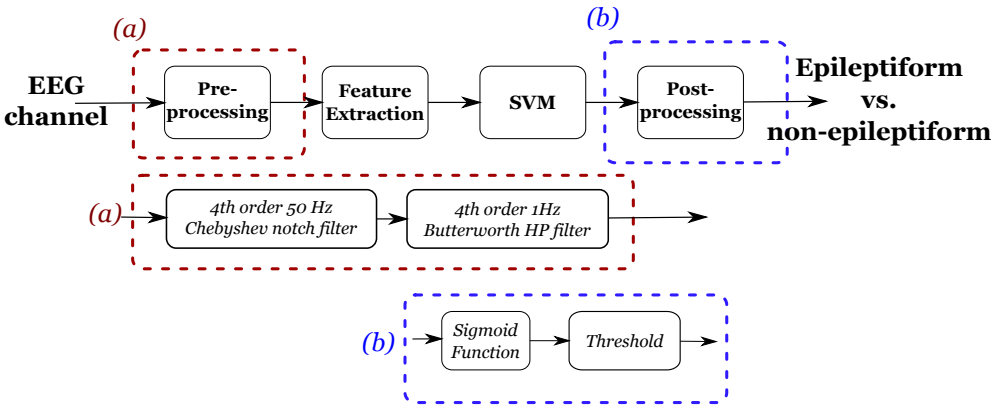


Figure 6.6: Overview of the epileptiform activity detection classifier.

the original (baseline) epileptic seizure and epileptiform activity classification system is illustrated in Figure 6.6. An outline of the classification architectures are displayed in Figure 6.7. The following paragraphs explain each of their stages in more detail.

6.5.2 Preprocessing

A fourth order Chebyshev notch filter was applied to the raw EEG data in order to remove the 50 Hz mains component from the EEG. An additional fourth order, Butterworth highpass filter with cut-off at 1 Hz was applied to remove the DC-component. The EEG was segmented into overlapping windows/epochs of 1 second. In order to maximise the amount of data available, overlap between windows of 90% was chosen (Kelleher et al., 2010).

6.5.3 EEG feature generation

A set of 55 features was extracted from the pre-processed EEG signal; the complete list of the features examined are displayed in Table 6.1. Feature selection was performed by Temko et al. (2011a) and showed that this feature set provides separation between neonatal seizure activity and normal EEG, and also between adult epileptic seizure activity and normal EEG (Faul, 2007; Faul et al., 2009). In Chapter 3 this feature set, supplemented with additional artefact specific features, was demonstrated to effectively separate between normal EEG and artefacts generated by head-movements (O'Regan et al., 2010a, 2013a). It could be argued that the results presented in Chapter 3 were over-optimistic in that background EEG and high-energy artefact signals may be easier to separate than

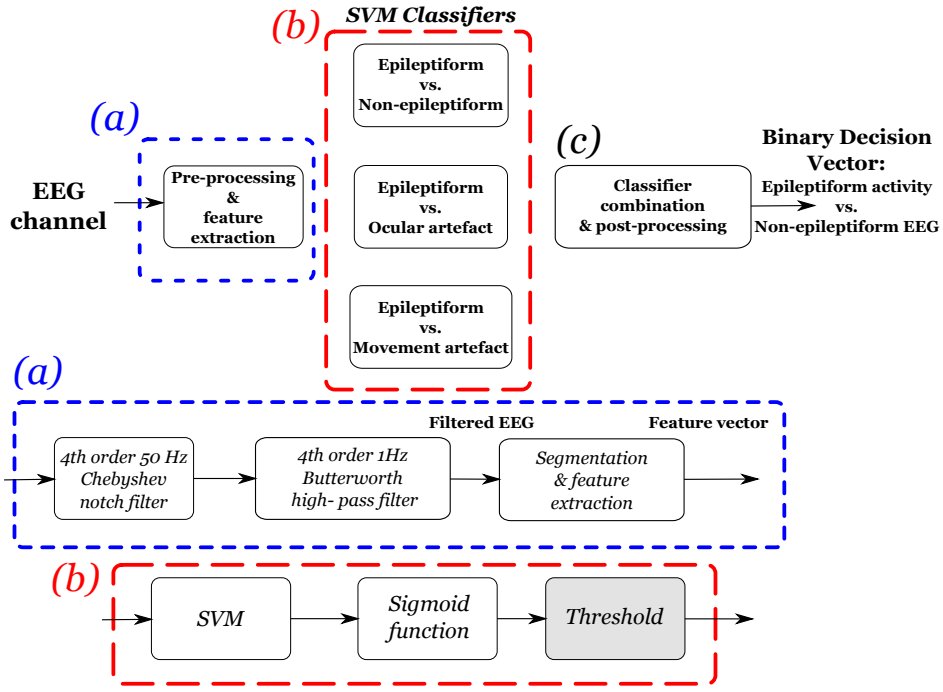


Figure 6.7: Overview of the classifier combination architectures. In block (a), filtering, segmentation and feature extraction is performed on the EEG. In block (b), features are independently presented to three different SVM models. In block (c) the outputs from the SVMs are combined. Note that the threshold (shaded) is applied in block (b) for decision fusion, but in block (c) for score fusion.

artefact and some neurological event such as epileptiform spikes. However, by examining the distributions of the features listed in Table 6.1 it can be seen that there is in fact considerable discrimination between epileptiform activity and artefact EEG. To illustrate this fact, probability density functions showing distributions of epileptiform, movement and ocular artefact EEG are displayed for a number of features in Figures 6.8, 6.9 and 6.10.

Probability density functions of line length (Equation 3.1) for epileptiform and ocular and movement artefacts are displayed in Figure 6.8. Here it can be seen that ocular artefacts are distributed largely at lower values compared to epileptiform or movement artefact EEG. Similarly, the ocular artefact EEG has a considerably narrower distribution than either epileptiform or movement artefact EEG, indicating that ocular artefacts tend to exhibit more homogeneous morphologies than either epileptiform or movement artefacts. Perhaps this is unsurprising, considering that ocular artefacts are usually easy to visually identify by their distinct shape, whereas both epileptiform and movement artefact can present in a wide variety of shapes and sizes. In the case of epileptiform this can include slow, sharp and slow and sharp wave complexes. Movement artefacts can comprise electrode pop, baseline drifts, and EMG artefact.

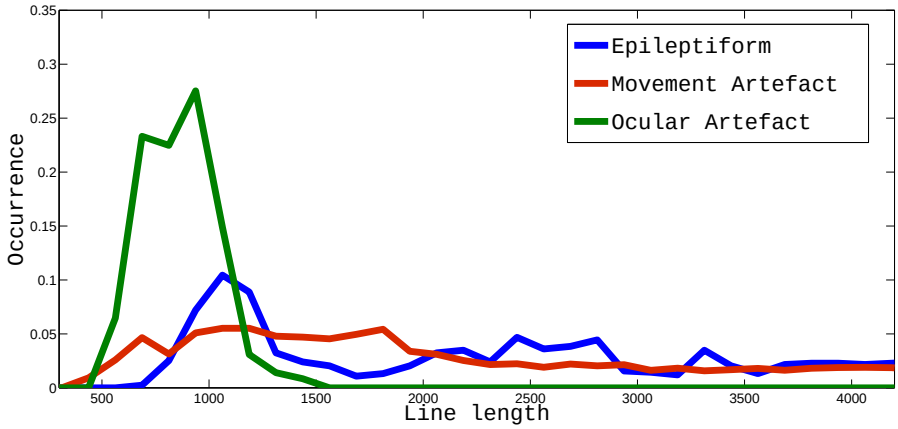


Figure 6.8: Probability density functions of line length for epileptiform EEG (blue) and ocular (green) and movement artefact (red).

Probability density functions of the Power in the 1-3 Hz frequency band for epileptiform and ocular and movement artefacts are displayed in Figure 6.9. As was the case for line length, the movement artefact EEG is almost uniformly distributed, again explainable by the fact that movement artefacts comprise EEG signals with large baseline shifts (high Power in this frequency band) as well as high-frequency EMG artefact (low Power in this

frequency range). The epileptiform EEG is broadly distributed with a peak for Powers of around 2.5 μW, indicating that in this frequency range, epochs with high Power are likely to be attributable to epileptiform activity. The Power of ocular artefact in this frequency range is lower than either epileptiform or movement artefact EEG.

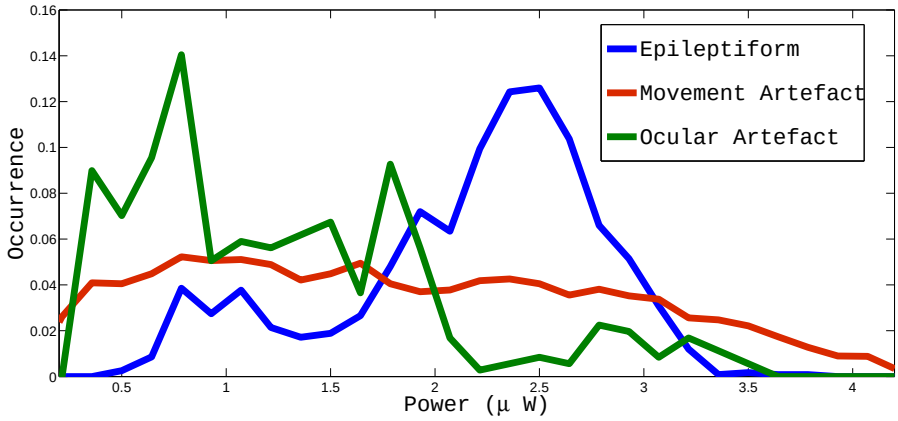


Figure 6.9: Probability density functions of the Power in the 1-3 Hz frequency band for epileptiform EEG (blue) and ocular (green) and movement artefact (red).

Figure 6.10 displays pdfs for the number of zero crossings for each of epileptiform EEG and movement and ocular artefact EEG. For this feature ocular artefacts typically have lower values; this is expected as both vertical and lateral movements generally result in a reasonably large baseline jump returning to the original amplitude after several hundred milliseconds (see Figure 6.5) and do not result in much oscillation about the origin. Movement artefact is more broadly distributed than ocular artefacts; the majority of movement artefacts result in relatively few zero crossings, explainable by the large baseline jumps of many movement artefacts. However, movement can also result in small amplitude EMG artefact, taking the form of high frequency oscillations that do not deviate much from the baseline axis; these artefacts then make up the high numbers of zero crossings in the distribution. The distribution of epileptiform activity is more uniform than both ocular and movement artefacts; in effect then, lower numbers of zero crossings are likely to indicate artefact, whereas larger numbers are likely to be epileptiform.

6.5.4 Classifiers

There are three SVM classifiers implemented in this experiment: the baseline classifier that flags epileptiform events (epileptiform activity vs. non-epileptiform activity), and two

Table 6.1: List of features for epileptiform activity detection.

Time Domain Features
- Line length/ Curve length(L)
- RMS Amplitude
- Activity (1st Hjorth Parameter)
- Mobility (2nd Hjorth Parameter)
- Complexity (3rd Hjorth Parameter)
- Kurtosis
- Skewness
- Nonlinear Energy (NLE)
- Zero Crossings (Zc)
- No. of inactive samples
- Autoregressive Model Error (AR models 1-9)
- Variance of first derivative
- Variance of second derivative
- Zero Crossings of first derivative
- Zero Crossings of second derivative

Frequency Domain Features
- Dominant Spectral Peak Frequency
- Wavelet Coefficients
- Total Spectral Power
- Spectral Edge Frequency(80, 90 and 95)
- Power Bands: 0-2 Hz, 1-3 Hz, 2-4 Hz, 3-5 Hz, 4-6 Hz, 5-7 Hz, 6-8 Hz, 7-9 Hz, 8-10 Hz, 9-11 Hz, 10-12 Hz
- Normalised Power Bands: 0-2 Hz, 1-3 Hz, 2-4 Hz, 3-5 Hz, 4-6 Hz, 5-7 Hz, 6-8 Hz, 7-9 Hz, 8-10 Hz, 9-11 Hz, 10-12 Hz

Information Theory-based Features
- SVD Entropy
- Shannon Entropy
- Fisher Information
- Spectral Entropy

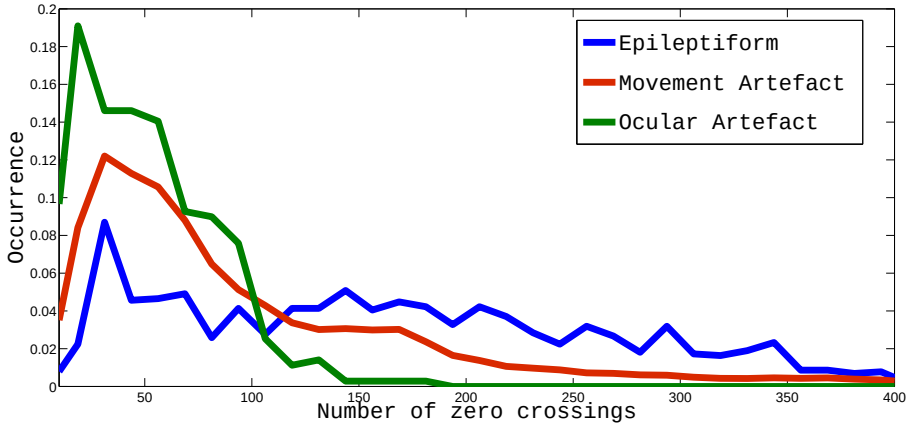


Figure 6.10: Probability density functions of the number of zero crossings for epileptiform EEG (blue) and ocular (green) and movement artefact (red).

artefact classifiers to differentiate between correctly identified epileptiform events and false positive events due to artefact (epileptiform activity vs. ocular artefact and epileptiform vs. movement artefact). In each classifier, the same pre-processing stage and feature set are used. A Gaussian kernel SVM is trained using 5,000 training examples from each class. Each classifier outputs the probabilities that the epoch in question contains epileptiform activity. In each of the three classifiers, the ratio of epileptiform to the other class (non-epileptiform, movement artefact or ocular artefact) was kept consistent. This makes possible the combination of classifier output probabilities as discussed in Section 6.6.1 (Kuncheva, 2004). An overview of the classification systems is displayed in Figure 6.7.

Epileptiform activity vs. non-epileptiform activity

In the baseline classifier, a Gaussian kernel SVM classifier was trained to separate between epileptiform activity (comprising epileptic seizure and short-duration, interictal, epileptiform activity) and non-epileptiform activity (comprising normal EEG and artefact) as outlined by Kelleher et al. (2010).

Epileptiform activity vs. movement artefact

This artefact classifier was designed to separate between epileptiform activity (comprising epileptic seizure and short-duration, interictal, epileptiform activity) and movement

artefacts.

Epileptiform activity vs. ocular artefact

This classifier was designed to separate between epileptiform activity (comprising epileptic seizure and short-duration, interictal, epileptiform activity) and ocular artefacts. The classifier is only applied to the four frontal EEG channels (FP2-F4, FP2-F8, FP1-F3 and FP1-F7), where ocular artefact is most pronounced.

6.5.5 Threshold choice

A binary output decision is obtained from each of the SVMs by applying a threshold to the classifier output probabilities; where a probability above or equal to the threshold is classified as epileptiform, and any probability below the threshold is considered to be non-epileptiform. As the system described here is designed to detect epileptic seizure as well as the shorter duration, interictal epileptiform activity, a suitable threshold θ must be chosen to ensure that all (or almost all) events are detected. However, the same leave-one-out cross-validation estimate for both threshold selection and performance evaluation cannot be used as this would introduce a (possibly quite strong) selection bias. A nested leave-one-out cross-validation procedure is therefore used instead. Leave-one-out cross-validation is used for performance evaluation in the "outer loop" of the procedure, in each iteration of which threshold selection is performed individually for each classifier based on a separate leave-one-out cross-validation procedure. Within each of these iterations, the threshold chosen was that maximum threshold for which all epileptiform events were correctly identified.

6.5.6 Metrics

As discussed in Section 2.9.4, ROC area and ROC_{sens95} area are used to assess the performance of the epileptiform detection systems outlined in this chapter. In addition to these epoch-based metrics, an event-based metric is also used. Traditionally, False Detections per hour (FD/h), an event-based metric, is used in conjunction with GDR. However, when artefact detection classifiers are fused with the baseline epileptiform classifier, the use of FD/h can lead to misrepresentative results. This occurs when a false detection, as

identified by the baseline detector is broken up into several independent events of shorter duration when the artefact detector is applied. This is illustrated in the right half of Figure 6.11, where the baseline classifier misinterprets a section of artefact EEG as epileptiform activity. When the epileptiform vs. ocular artefact detection classifier is combined with the baseline classifier, some of this output decision is corrected. However, this leads to an increase in FD/h as there are now two false positive events detected instead of one. For this reason, FDR is used to quantify the false detections in each classifier architecture.

Consequently, *False Detection Rate (FDR)*, another epoch-based metric, is used in conjunction with GDR in assessing the performance of different classifier architectures in this study. False detection rate is the percentage of non-epileptiform epochs (comprising normal EEG and artefact) which are falsely classified as representing epileptiform activity, and defined in Equation 6.1 as:

$$\text{FDR} = 1 - \textit{Specificity} \tag{6.1}$$

As discussed in Section 6.5.5, the threshold θ is chosen via nested LOO-CV. This results in two FDR measures: the FDR achieved on the "inner loop" of the nested LOO-CV, FDR_{tr} and the FDR on the evaluation (test) data, FDR_{te}.

6.6 Classifier fusion architectures

The objective of the classifier combinations outlined in this chapter is to reduce the number of false positive detections by the epileptiform activity detection classifier while continuing to correctly detect all (or almost all) epileptiform events. In contrast to the classifier fusion methods outlined in Chapter 5, classifier outputs are combined only on those epochs that are flagged by the baseline epileptiform classifier as containing epileptiform activity. Thus, any epochs that have been flagged by the baseline epileptiform classifier (epileptiform vs. non-epileptiform) are combined with the corresponding epoch and channel of the artefact classifier (epileptiform vs. ocular, epileptiform vs. movement, or both). As discussed in Chapter 5, classifier combination methods can be sub-divided into two broad categories: combination of the classifier output probabilities (score fusion) and combination of the classifier output decisions (decision fusion). In this chapter, the best-performing score and decision fusion methods from Chapter 5 are investigated for the epileptiform activity detection task.

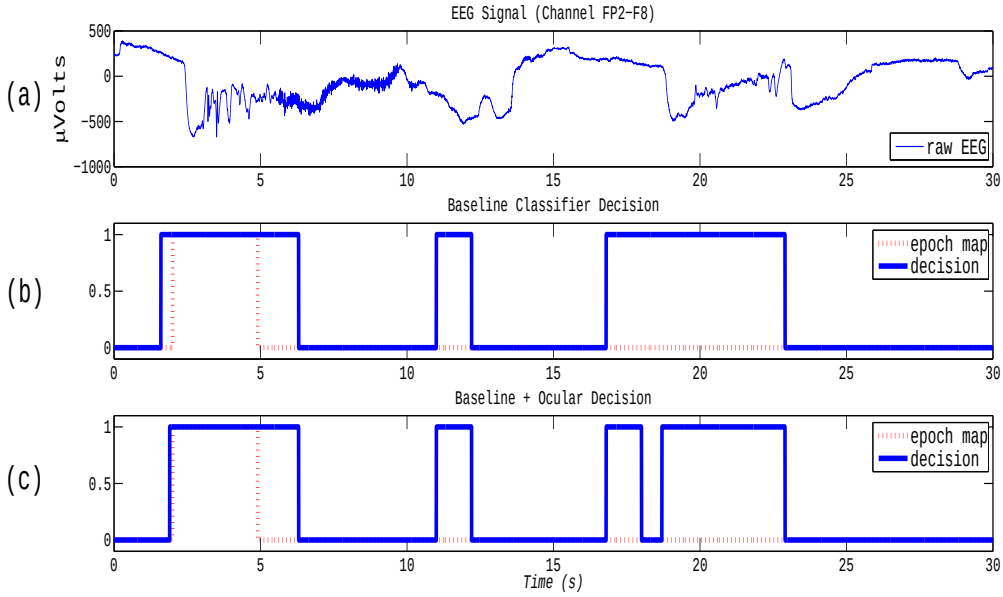


Figure 6.11: Illustration of classifier output decisions. (a) displays a thirty second section of EEG taken from channel FP2-F8. (b) shows the baseline epileptiform classifier binary decision for the EEG (blue) and annotations (red). (c) displays shows the output decision for the baseline and ocular classifiers deployed together (blue) and the epileptiform annotations (red).

6.6.1 Score fusion

As outlined in Chapter 5, score fusion combines classifier output scores; in this case, the posterior probabilities from SVM after the sigmoid function has been applied. In the score-level classifier fusion performed in this experiment, any epochs that have been flagged by the baseline epileptiform classifier (epileptiform vs. non-epileptiform) are combined with the corresponding epoch and channel of the artefact classifier. This is equivalent to taking the mean of the probability outputs from each SVM for those epochs that are flagged as epileptiform by the baseline classifier. Taking the mean of the classifier output probabilities has been shown to be the most resilient to estimation errors and as such is usually the best-performing fixed-rule, score level classifier fusion method (Kittler et al., 1998). In Chapter 5, it was demonstrated that this was indeed the case in combining EEG and gyroscope signals in detecting head-movement artefacts. This method is described for a single artefact classifier in Algorithm 1 and for dual artefact classifiers in Algorithm 2, where P_E is the output score from the baseline epileptiform vs. non-epileptiform classifier, P_{E-MOV} is the output score from the epileptiform vs. movement classifier and P_{E-OC} is the output score from the epileptiform vs. ocular classifier.

The behaviour of this fusion method is best explained by means of an example. If the baseline epileptiform classifier finds that an epoch has a strong probability of containing epileptiform activity, and the epileptiform vs. artefact classifier, similarly agrees that the epoch likely contains epileptiform activity, the final probability output will reflect this. Conversely, in the case of a false positive epileptiform event where the baseline epileptiform classifier flags the epoch as containing epileptiform activity (as the probability output $P(E|x)$ is greater than the set threshold θ), the epileptiform vs. artefact classifier should predict that there is a low chance that the epoch contains epileptiform activity, thus reducing the final probability output.

Algorithm 1: Score fusion algorithm for applying one artefact detector in combination with the baseline epileptiform classifier. In this example, the movement artefact detector is applied.

```
if P_E > θ then
    if P_E + P_{E−MOV} > 2θ then
        Class → Epileptiform;
    else
        Class → Non-Epileptiform;
    end
else
    Class → Non-Epileptiform;
end
```

Algorithm 2: Score fusion algorithm when both movement and ocular artefact detection classifiers are combined with the baseline epileptiform classifier.

```
if P_E > θ then
    if P_E + P_{E−OC} + P_{E−MOV} > 3θ then
        Class → Epileptiform;
    else
        Class → Non-Epileptiform;
    end
else
    Class → Non-Epileptiform;
end
```

6.6.2 Decision fusion

Classifier fusion at the decision level is performed on the binary classifier output decisions, once a threshold has been applied to the classifier posterior probabilities. All epochs from

the main epileptiform classifier that are flagged as epileptiform in origin are combined with the corresponding epochs from the artefact classifier. The decision fusion is performed using the logical AND operation (if 1 exists in the output decision of the baseline epileptiform classifier and in the output decision of the artefact detection classifier, then the result of the combination of the two decision vectors will also be a 1). This process is illustrated for a single artefact classifier in Algorithm 3 and for dual artefact classifiers in Algorithm 4, where P_E is the binary decision output from the baseline epileptiform vs. non-epileptiform classifier, P_{E-MOV} is the binary output decision from the epileptiform vs. movement classifier and P_{E-OC} is the binary output decision from the epileptiform vs. ocular classifier.

In both the score and decision fusion methods, once the classifiers have been combined, the sum of decisions across channels is taken and a single decision for the presence or otherwise of an epileptiform event in the current epoch is obtained, i.e. if an event is detected in one or more channels then the current epoch is labelled as containing epileptiform activity.

Algorithm 3: Decision fusion algorithm for applying one artefact detector in combination with the baseline epileptiform classifier. In this example, the movement artefact detector is applied.

if $P_E > \theta \quad AND \quad P_{E-MOV} > \theta$ **then**
  Class $\rightarrow$ Epileptiform;
else
  Class $\rightarrow$ Non-Epileptiform;
end

Algorithm 4: Decision fusion algorithm when both movement and ocular artefact detection classifiers are combined with the baseline epileptiform classifier.

if $P_E > \theta \quad AND \quad P_{E-OC} > \theta \quad AND \quad P_{E-MOV} > \theta$ **then**
  Class $\rightarrow$ Epileptiform;
else
  Class $\rightarrow$ Non-Epileptiform;
end

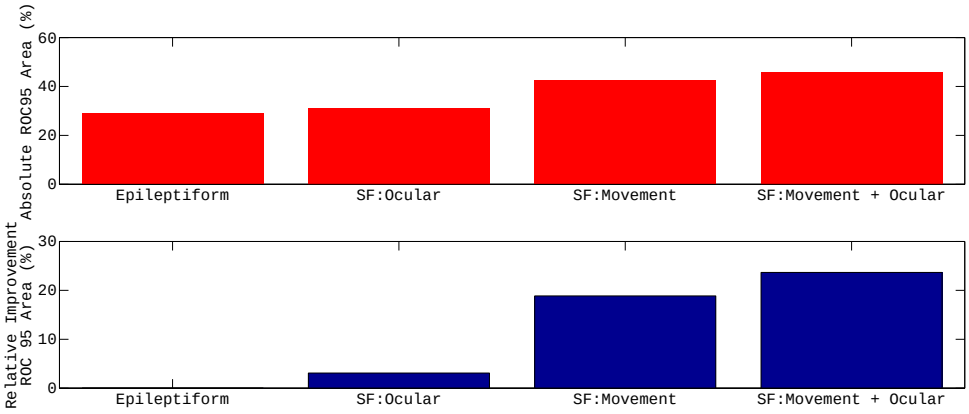


Figure 6.12: Mean ROC 95 areas (red) and the relative improvement in mean ROC_{sens95} area (blue) for the baseline epileptiform classifier and each score fusion (SF) classifier combination. Random discrimination gives a ROC 95 area of 2.5%.

6.7 Comparison of fusion methods

6.7.1 Epoch-based metrics

As outlined in Section 2.9.4, ROC_{sens95} area gives an indication of the classifier performance over a range of threshold values, thus allowing the user to specify an operating point at which the system performs as desired. Figure 6.12, displays the ROC_{sens95} areas and relative improvements in ROC_{sens95} areas for each of the score fusion classifier architectures. The ROC_{sens95} shows a relative improvement of 3.1 % when the epileptiform vs. ocular artefact classifier is combined with the baseline epileptiform detector. When the epileptiform vs. movement artefact classifier is applied, there is a relative improvement of 18.87 %. The greatest relative improvement (23.66 %) is seen when both the ocular and movement artefact detection classifiers are combined simultaneously with the baseline epileptiform classifier. The greater relative improvement of classifier performance by applying the movement artefact detector can be largely attributed to the fact that movement artefacts accounted for $\sim$ 40% of original false detections compared to $\sim$ 20% false detections due to ocular artefacts as well as the fact that the ocular artefact detector was applied to frontal EEG channels only.

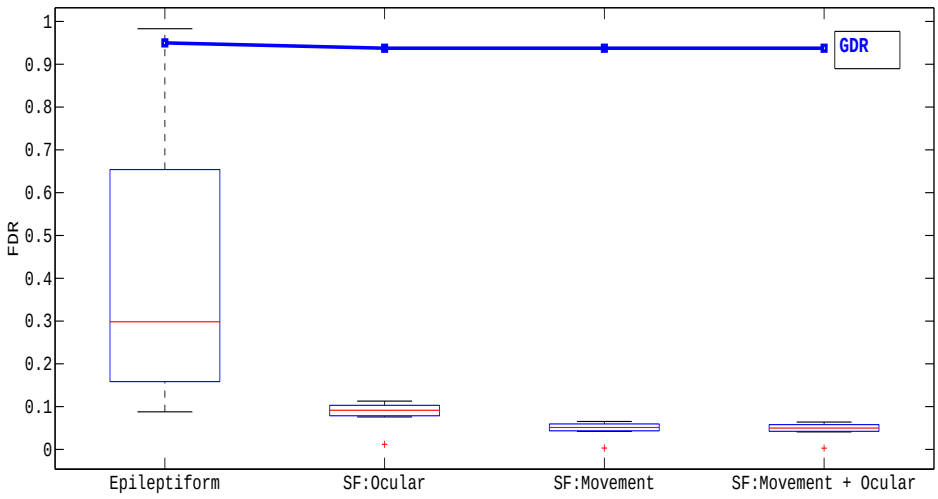


Figure 6.13: Boxplots of FDR for the baseline epileptiform classifier and each of the score fusion (SF) classifier architectures. The corresponding mean GDRs are plotted in bold above.

6.7.2 Event-based metrics

As discussed in Section 2.9.4, it is necessary in practice to specify a single threshold θ at which the classification decision will be made. Figure 6.13 shows boxplots of the FDR_{te} for the baseline epileptiform classifier and each of the score fusion classifier architectures. It is evident from the graph that a substantial improvement in FDR_{te} is seen when each of the artefact detection classifiers is applied. The median FDR_{te} drops by 69.28 % when the ocular artefact detection classifier is combined with the baseline epileptiform classifier. When the movement artefact detection classifier is used, median FDR_{te} drops by 82.76 %. The greatest improvement in performance is seen when both ocular and movement classifiers are applied simultaneously, resulting in the median FDR_{te} falling by 83.27 %. These reductions in FDR_{te} are achieved while mean GDR remains largely constant. A further notable trend in the results is that the variability of FDR_{te} values narrow considerably when the artefact classifiers are applied. This indicates that the introduction of artefact-specific processing on sections of suspected epileptiform activity adds an element of robustness to the classification task, with the epileptiform activity and epileptic seizure detection system performing similarly for all patients. This has considerable benefits in a clinical setting as it ensures that patients recordings will produce similar numbers of false positive detections, allowing the clinician to take this into account.

Figure 6.14 displays the FDR_{te} results for the movement and ocular artefact classifiers applied simultaneously in combination with the baseline epileptiform classifier for both score and decision fusion classifier architectures. There is a clear disparity between the two fusion methods which can be explained by the fact that the score fusion method of combining classifiers offers more flexibility in threshold choice. As the same θ value is applied to each classifier, in the decision fusion architecture the classification performance is likely to be hindered by its worst performing constituent classifier.

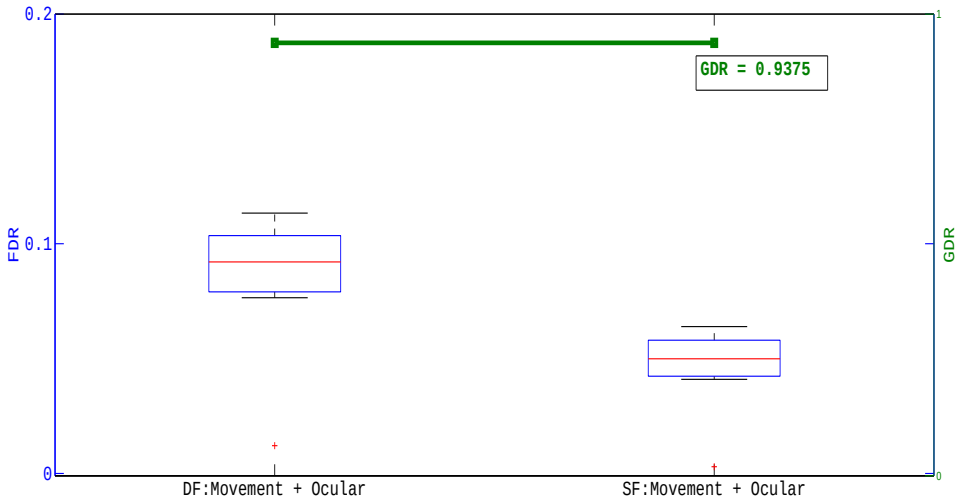


Figure 6.14: Score (SF) and decision (DF) fusion boxplots of FDR for movement and ocular artefact detection classifiers applied simultaneously in combination with the baseline epileptiform classifier.

The validity of the threshold selection process is highlighted in Figure 6.15, which shows FDR_{tr} and FDR_{te} for both artefact classifiers applied in combination with the baseline epileptiform classifier using the score fusion method. It is expected that the training results should be better than the testing results and this is reflected by the fact that there is a drop in the GDR for the testing results. Additionally, the variability of the FDR_{tr} values is substantially smaller than that of the FDR_{te} values. Figure 6.15 also includes the test results for the "oracle" system. The "oracle" is defined here as the performance that would be achieved if perfect knowledge of the test data was available when the threshold θ was chosen and defines a lower bound on the FDR. In other words, with the a classifier tested on unseen data, the "oracle" defines the performance obtained by the optimally chosen threshold and is in effect the best performing point on the ROC curve. It is evident that FDR_{te} approaches the "oracle" FDR performance, albeit at a drop in GDR.

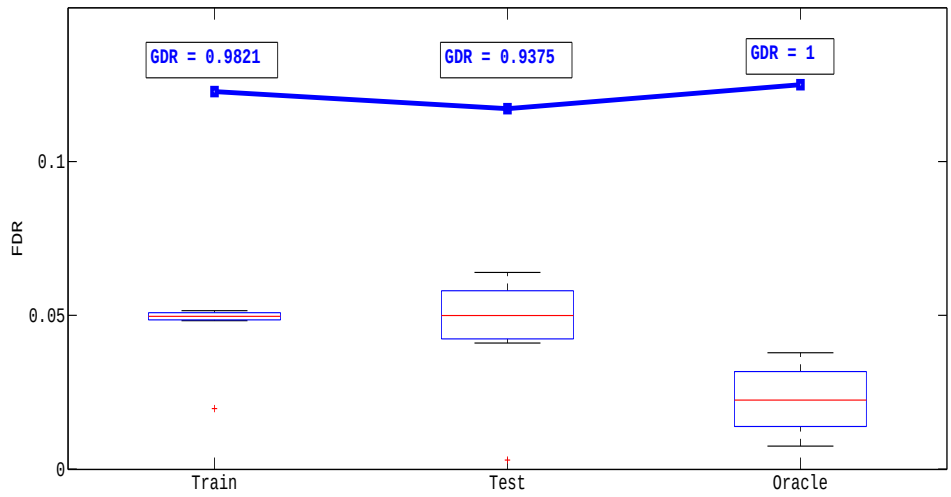


Figure 6.15: Boxplots of FDR for train, test and oracle thresholds. The corresponding mean GDRs are plotted in bold above.

6.8 Summary

Using both epoch and event-based metrics, it is clear that the use of artefact-specific SVM classifiers in EEG sections of suspected epileptiform activity improves the performance of the automated epileptiform detection system. The FDR metric verifies that a considerable reduction in the number of false detections is achieved whilst continuing to accurately detect epileptic seizure and short-duration, interictal epileptiform events; median false detection rate drops by 83.27 % at a constant good detection rate of 0.9375. This result underlines the performance gain achieved by including artefact-specific classifiers and fusing them at the post-processing stage. The most profound improvement is found by incorporating head-movement and ocular artefact detection classifiers together and combining them with the baseline epileptic seizure and epileptiform activity detection classifier. The performance gains in using each of the artefact classifiers were seen across patients. This feature is particularly welcomed as it adds a level of robustness to the epileptiform detection algorithm that is necessary in a real-world setting. In light of these results, the use of score level classifier fusion is recommended as it offers superior performance to that of the decision level classifier fusion, reducing median false detection rate from 0.3 to 0.05.

Chapter 7

Artefact removal for automated neonatal seizure detection

7.1 Introduction

Artefacts are a major obstacle in the deployment of automated neonatal seizure detection algorithms in the NICU, as they lead to large numbers of false positive detections. In babies with suspected seizure, persistent false alarms may prompt the on-duty nursing staff to abandon the seizure detection algorithm. More critically, false detections are unacceptable as they may lead to patients being incorrectly treated for seizure (by cooling or with anti-seizure medication). Amongst the array of artefact types, respiratory artefacts in particular have proven problematic for automated neonatal seizure detection algorithms as they are often morphologically similar to seizure. Indeed, the problem of this similarity is not limited to automated systems; respiration artefact has been widely acknowledged as being troublesome for human experts with years of experience. In the previous chapter, emphasis was placed on the use of supervised machine learning algorithms to identify artefacts; in this chapter, artefact processing follows the well-established research area of artefact removal using blind source separation methods to "clean" the EEG signal that has been contaminated by these non-cerebral artefacts.

From a utilitarian perspective, artefact removal is only of interest if it improves the performance of the seizure detection classifier. Therefore, in this chapter, several respiratory artefact removal methods will be evaluated in the context of neonatal seizure detection.

These methods will combine automated blind source separation of the EEG with additional physiological signals in a pre-processing stage in the automated neonatal seizure detection algorithm. In doing so, effective means of removing respiratory artefact in neonatal EEG are developed, leading to improvement in the neonatal seizure detection algorithm.

One journal publication (O'Regan et al., 2013c) is in preparation based upon the work carried out in this chapter.

7.2 Neonatal seizure detection

7.2.1 Neonatal seizure

As discussed in Chapter 2, the most prominent feature of neurological dysfunction in the neonatal period is the occurrence of seizures. These seizures are clinically defined as paroxysmal alterations in neurological function, i.e. an alteration in behavioural, motor and/or autonomic function and are powerful predictors of long-term cognitive and developmental impairment (Volpe, 1989). The EEG is considered the gold standard in neonatal seizure diagnosis with seizures manifesting as repetitive patterns with minimum duration of 10 seconds (Clancy, 2006).

However, the diagnosis of electrographic seizure is non-trivial; the appearance of the discharge can transform in frequency, amplitude and morphology over time (Mizrahi and Clancy, 2000). Electrographic seizures are predominantly unifocal but multifocal seizures may also occur in different brain regions and fire simultaneously and asynchronously. Both unifocal and multifocal seizure discharges can spread from one location via abrupt change or by gradual widening. Examples of seizure are illustrated in Figures 7.1 and 7.2. Figure 7.1 displays a generalized neonatal seizure with seizure activity present on all channels and is most prominent on channels C4-O2, C3-O1 and C3-T3 (red). Figure 7.2 displays a localised neonatal seizure with slow baseline oscillations visible on channels F3-C3, C3-01, Cz-C3 and C3-T3 (red).

7.2.2 Automated neonatal seizure detection

Few staff members of the neonatal intensive care unit receive sufficient training to interpret EEG traces. Automated neonatal seizure detectors have thus been proposed to play an

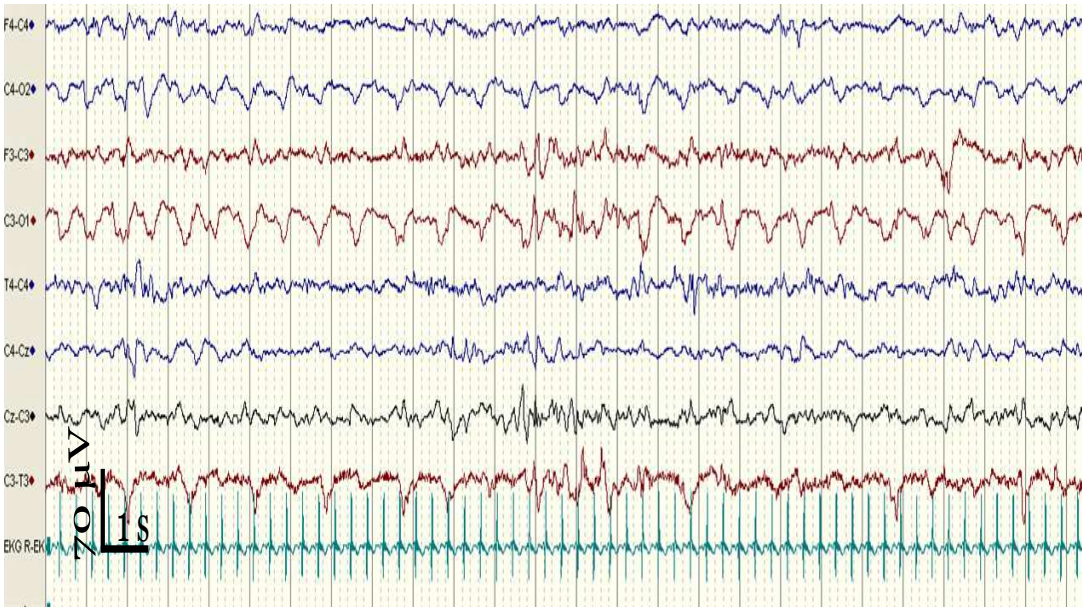


Figure 7.1: Example of neonatal EEG showing a generalised neonatal seizure. Seizure is present on all channels but is most prominent on channels C4-O2, C3-O1 and C3-T3 (red).

assistive role in communicating the occurrence of seizure to on-duty nursing staff. The ANSeR seizure detection system developed over the past decade at U.C.C. is outlined in Figure 7.3. The preprocessing, feature extraction and classification sections are performed on a per-channel basis, similar to the epileptiform activity detection system outlined in Chapter 6. The spikiness of SVM output probabilities is then smoothed with a moving average filter and a threshold is applied to obtain a binary output decision for each EEG channel. If a seizure is detected on any channel, the ANSeR system flags the seizure and uses a collar operation (collar width of 40 seconds) to identify surrounding epochs as also containing seizure (Temko et al., 2011b).

7.3 Respiratory artefact and false detections

Despite state-of-the-art performance of the SDA described in Section 7.2.2, the seizure detection system is nevertheless prone to false detections. In seeking to explain the underlying causes of these false detections, Thomas (2011) has pointed out that neonatal seizure is composed of a large set of diverse patterns emanating from background activity, seizure events and artefacts. Additionally, EEG characteristics are often considerably different between patients, and are not stationary, evolving over time, particularly in the immediate weeks after birth. Perhaps the most critical finding with respect to artefact

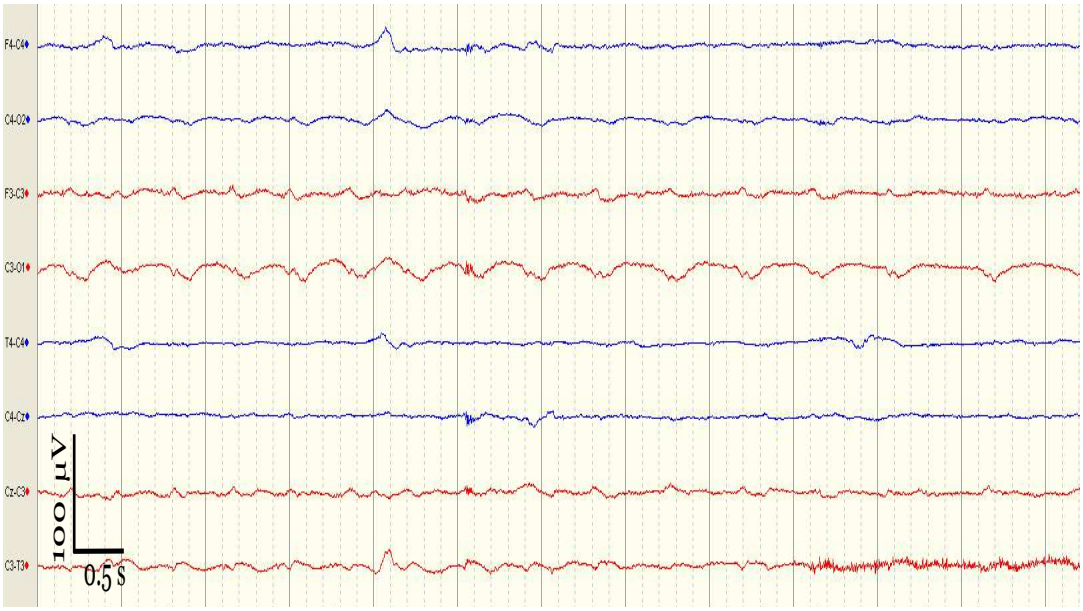


Figure 7.2: Example of neonatal EEG showing a localised neonatal seizure. Seizure is visible on channels F3-C3, C3-01, Cz-C3 and C3-T3 (red).

by Thomas (2011) was that artefacts accounted for 43% of false positive detections. In a subsequent clinical evaluation of the SDA, Mathieson (2012) confirmed that contamination of the neonatal EEG by non-cerebral electrical activity is the single largest cause of misclassification in the ANSeR system developed by Faul (2007), Thomas (2011) and Temko et al. (2011b).

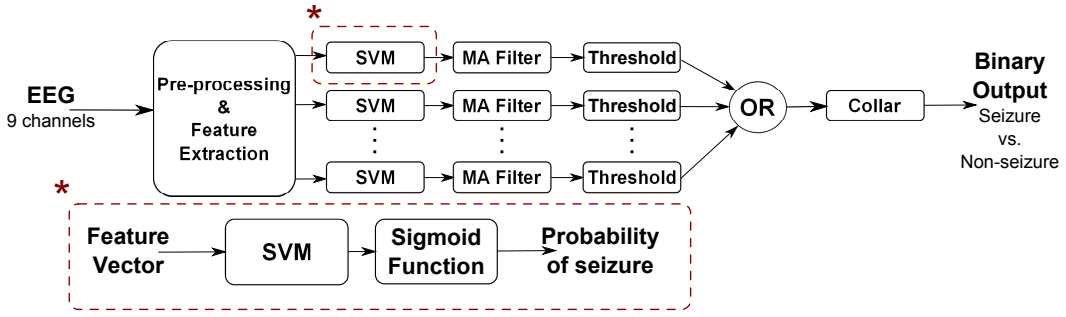


Figure 7.3: Illustration of the ANSeR neonatal seizure detection algorithm developed over the last ten years by the Neonatal Brain Research Group at U.C.C.

In the investigation by Mathieson (2012), it was clear that certain artefact types were more problematic than others in the context of SDA performance. Most detrimental to the SDA performance was contamination by respiratory artefacts (Table 7.1), i.e. artefacts arising from the breathing of the neonate under observation. As discussed in Chapter 2, respiration artefacts occur due to movement of an electrode with inhalation or exhalation.

SDA Threshold	FD\h	FDs due to respiratory artefact	Percentage FDs due to respiratory artefact
0.4	0.63	278	34.7 %
0.5	0.41	249	47.8 %
0.6	0.27	221	64.61 %

Table 7.1: Breakdown of false detections by the ANSeR seizure detection algorithm due to respiratory artefact as reported in Mathieson (2012).

This can manifest itself in one of two ways: slow waves or slow and sharp waves. Slow waves are respiration artefacts taking the form of slow, rhythmic EEG activity, synchronous with the body movements associated with breathing and mechanically affecting the impedance of (usually) one electrode. A second form of respiration artefact, which is especially common in neonatal EEG, can be slow or sharp waves that occur synchronously with inhalation or exhalation and involve those electrodes upon which the patient is lying. This form of respiration artefact can often mimic neonatal seizure, and is understood to play the largest role in the false detections observed by the SDA in the presence of respiratory artefacts. Examples of respiration artefact are illustrated in Figures 7.4 and 7.3. The likeness of these respiration artefacts to seizure can be seen with comparison to the seizure examples illustrated in Figures 7.1 and 7.2, where the similarity of their morphologies can be observed.

7.4 Automated artefact removal

As described in Chapters 2, artefact removal techniques can be broadly split into two categories: (i) filtering and regression and (ii) blind source separation. Generally speaking, removal of non-physiological artefacts is often straightforward as these artefacts typically occupy well-defined frequency and amplitude ranges. Thus, widely-accepted methods of removing non-physiological artefacts are linear filtering and simple amplitude thresholding. In the ANSeR system outlined in Section 7.2.2, linear filters are included in the pre-processing stage to remove DC and mains frequency noise from the EEG. However, for physiological artefacts such as respiratory artefact, more advanced signal processing techniques are required; most commonly blind source separation techniques are used.

As discussed in Chapter 2, blind source separation uses information from multivariate EEG to construct more insightful new variables by applying a coordinate transformation to the data. In this new reference frame, the artefact components are identified and

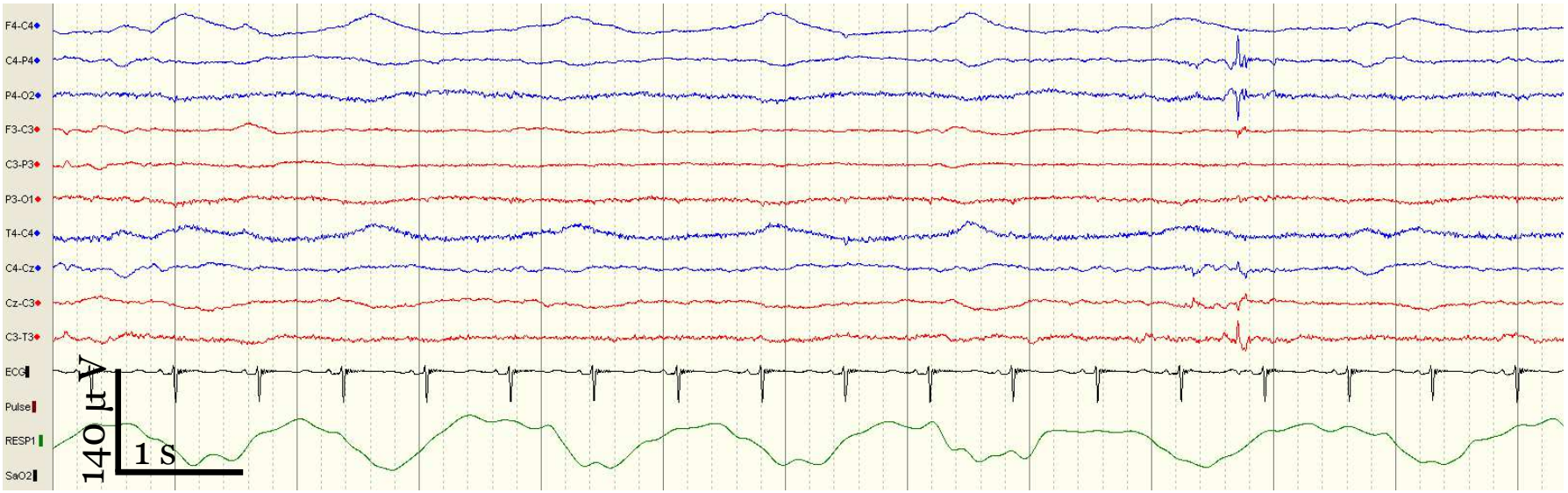


Figure 7.4: Example of respiration artefact in EEG recorded from neonatal patients at CUMH using a NicOne EEG machine using a sampling frequency of 250 Hz. Respiration artefact, manifesting as slow waves, on neonatal EEG recording; electrode C4 records the artefact which is then visible on channels F4-C4 and T4-C4, and to a lesser extent on channels C4-P4 and C4-Cz.

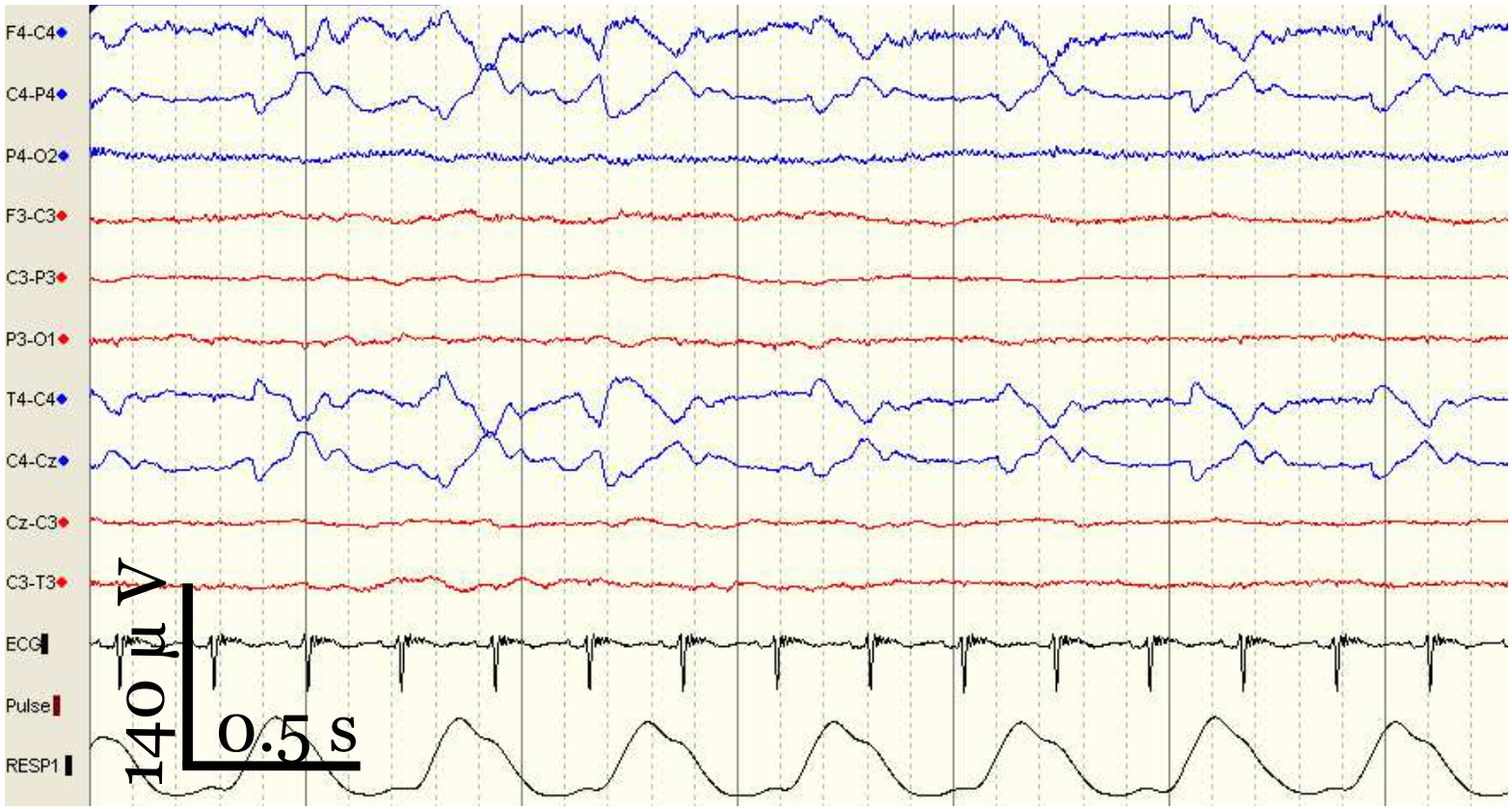


Figure 7.5: Example of respiration artefact in EEG recorded from neonatal patients at CUMH using a NicOne EEG machine using a sampling frequency of 250 Hz. Respiration artefact on neonatal EEG recording; visible on channels F4-C4, C4-O2, T4-C4 and C4-Cz, manifesting as slow and sharp waves.

removed before performing an inverse coordinate transformation back to the original reference, where the multivariate EEG has now been "cleaned". In the context of neonatal seizure detection, if detection of components containing artefact is too sensitive, too many sources may be omitted, resulting in data loss and possibly missed seizures. Conversely, if artefact identification is not sensitive enough, artefacts will remain in the EEG, resulting in continued false positive seizure detections. To date, there exists one method in the literature that proposes to effectively remove respiratory artefact from EEG in the context of automated neonatal seizure detection (De Vos et al., 2011). In De Vos et al. (2011), the authors outline an automated respiratory artefact removal algorithm utilising second order blind identification (SOBI) and a respiratory trace. In this chapter, the algorithm outlined by De Vos et al. (2011) is recreated and evaluated in the context of the ANSeR seizure detection system developed at U.C.C. This algorithm is then further developed to function without a respiratory trace.

7.5 Data

The dataset used in this chapter is outlined in detail in Section 2.9.3. To summarise, the ANSeR system was developed utilising data from 55 babies recorded at Cork University Hospital. The artefact removal algorithms outlined in this chapter are evaluated on neonatal data collected at University College London Hospitals; the data is taken from neonatal patients whose recordings show seizure as well as respiration artefact. The availability of data from neonatal patients who underwent seizure, whose EEG was contaminated by respiration artefact and where a respiration trace was recorded was limited to two patients. In this chapter, the De Vos algorithm is implemented with the ANSeR system with 36 hours of data recorded at University College London Hospitals from two neonatal patients (referred to here as Patients 1 and 2).

The second algorithm evaluated using the ANSeR system is a method developed in this chapter that adapts the De Vos algorithm for use with a respiration signal derived from the ECG signal. By relaxing the need for a respiration trace, two more babies can be recruited. Thus, this algorithm is evaluated on the aforementioned two patients as well as an additional two babies (Patients 3 and 4) that suffered from seizure, and whose EEG was contaminated by respiration artefact. These recordings were not accompanied by a respiration trace, but did contain ECG signals.

7.6 De Vos method: correlating SOBI with respiratory trace

In the respiratory artefact removal method outlined by De Vos et al. (2011), the EEG is first decomposed into its prominent underlying sources using an appropriate blind source separation method. It is then necessary to automatically identify which (if any) of these sources correspond to respiratory artefact. To accomplish this, each source is compared to a simultaneously recorded polygraphy signal, i.e. the respiration trace in the form of a movement sensor on the abdomen of the neonate. An illustration of the De Vos algorithm is outlined in Figure 7.6, with further details of the algorithm described in the following paragraphs.

7.6.1 Second order blind identification

As respiration artefacts tend to be oscillatory in nature and consequently have high autocorrelation, independent component analysis may be a suboptimal method of performing blind source separation. Second order blind identification (SOBI) introduced by Belouchrani et al. (1997) is an alternative BSS algorithm that is appropriate for separating sources that are individually correlated in time, but mutually uncorrelated. SOBI is based on a joint diagonalisation of correlation matrices, and considers the relationship between component values at different time lags and decorrelates these values as much as possible; thus, SOBI uses correlations across time in performing the signal separation. Consequently, this means that SOBI can isolate highly temporally correlated sources, something that most ICA algorithms cannot do (Joyce et al., 2004). For this reason, SOBI is proposed by De Vos et al. (2011) as a suitable means of performing source separation in the presence of oscillatory, respiration artefacts.

As discussed in Section 2.5, the recorded EEG signals $\mathbf{X}$, can be modelled as a linear transformation of latent variables $\mathbf{Y}$ comprising underlying cortical signals as well as electrical signals arising from non-cerebral sources, such that the transformation of the N-channel EEG is modelled by an unknown, full-rank, mixing matrix, $\mathbf{A}$:

$$\mathbf{X} = \mathbf{AY}, \tag{7.1}$$

where $\mathbf{A} \in \mathbb{R}^{N \times N}$. More commonly for time-series data such as EEG signals, the data is represented as:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{y}(t) + \mathbf{n}(t), \tag{7.2}$$

where $\mathbf{n}(t)$ is additive noise and is modelled by a stationary, temporally and spatially white zero-mean process with variance σ^2 and where $\mathbf{x}(t)$ denotes the signals and $\mathbf{y}(t)$ denoting the latent variables are sources. It is assumed that the n sources $\mathbf{y}(t)$ are mutually uncorrelated and that $m \times n$ complex matrix $\mathbf{A}$ is of full rank but otherwise unknown. The autocovariance of the sources is thus:

$$\mathbf{R}_y(\tau) := \mathbf{E}\left((\hat{\mathbf{y}}(t+\tau) - \mathbf{E}(\mathbf{y}(t)))\,(\mathbf{y}(t) - \mathbf{E}(\mathbf{y}(t)))^H\right), \tag{7.3}$$

for all τ and where superscript H denotes the complex conjugate transpose of the matrix. By centering the processes, it can be assumed that $\mathbf{x}(t)$ and hence $\mathbf{y}(t)$ have zero mean. The autocovariances then have the following structure:

$$\mathbf{R}_x(\tau) = \mathbf{E}\left(\mathbf{x}(t+\tau)\mathbf{x}(t)^H\right) = \begin{cases} \mathbf{A}\mathbf{R}_y(0)\mathbf{A}^H + \sigma^2\mathbf{I} & \tau = 0 \\ \mathbf{A}\mathbf{R}_y(\tau)\mathbf{A}^H & \tau \neq 0 \end{cases} \tag{7.4}$$

The goal of blind source separation is then to estimate an inverse of $\mathbf{A}$ that allows the accurate recovery of the sources $\mathbf{y}(t)$, without use of any *a priori* information on $\mathbf{A}$ and thus using only information from the signals $\mathbf{x}(t)$. In other words, find the transformation $\mathbf{A}^{-1}$ to coordinates corresponding to the underlying cortical signals as well as artefact sources. As this transformation is simply an estimate of the actual de-mixing matrix $\mathbf{A}^{-1}$, it will be referred to here as $\hat{\mathbf{A}}^{-1}$.

Clearly, $\mathbf{A}$ (and hence $\mathbf{y}(t)$) can be determined by Equation 7.2 only up to permutation and scaling of columns. Since existing variances of $\mathbf{x}(t)$ and hence $\mathbf{y}(t)$ are assumed, the scaling indeterminacy can be eliminated by the convention $\mathbf{R}_y(0) = \mathbf{I}$. In order to guarantee the identifiability of $\mathbf{A}$ (except for permutation) from the above model, it is necessary to additionally assume that there exists a delay τ such that $\mathbf{R}_y(\tau)$ has pairwise different eigenvalues (Belouchrani et al., 1997). Then using the spectral theorem (Halmos, 1963) it can be seen from Equation 7.4 that $\mathbf{A}$ is determined uniquely by $\mathbf{x}(t)$ except for permutation.

To recover $\mathbf{A}$, the no-noise EEG term $\hat{\mathbf{x}}(t) := \mathbf{A}\mathbf{y}(t)$ is first *whitened*, using an invertible matrix $\mathbf{W}$ such that $\mathbf{W}\hat{\mathbf{x}}(t)$ has unit covariance. The whitening matrix $\mathbf{W}$ is estimated by diagonalising the sample covariance matrix $\mathbf{R}_{\hat{\mathbf{x}}}(0)$. In addition to only diagonalizing a single autocovariance matrix, SOBI takes a whole set of autocovariance matrices of $\mathbf{x}(t)$ with varying time lags τ and jointly diagonalizes the whole set. That is to say, SOBI finds $\hat{\mathbf{A}}^{-1}$ by minimizing the sum squared cross-correlations between one component at time t

and another component at time $t+\tau$, across multiple time delays (τs) and across all pairs of components. The SOBI algorithm as detailed by Belouchrani et al. (1997) is thus:

1. Estimate the sample covariance $\hat{\mathbf{R}}(0)$ from T samples. Denote the largest n eigenvalues of $\hat{\mathbf{R}}(0)$ as $\lambda_1, ..., \lambda_n$ and their corresponding eigenvectors, $\mathbf{h}_1, ..., \mathbf{h}_n$.

2. Obtain the whitened signals $\mathbf{Z} = \hat{\mathbf{W}}\mathbf{X}$, where the whitening matrix $\hat{\mathbf{W}}$, is such that: $\hat{\mathbf{W}} = [(\lambda_1 - \hat{\sigma}^2)^{-\frac{1}{2}}\mathbf{h}_1, ..., (\lambda_n - \hat{\sigma}^2)^{-\frac{1}{2}}\mathbf{h}_n]^H$, where superscript H is the complex conjugate transpose, and $\hat{\sigma}^2$ is an estimate of the variance of the noise obtained by taking the average of the $m - n$ smallest eigenvalues of $\hat{\mathbf{R}}(0)$.

3. Form sample estimates $\hat{\mathbf{R}}_y(\tau_j)$ by computing the sample covariance matrices of the whitened signals $\mathbf{Z}$, for a fixed set of time lags $\tau \in \{\tau_j | j = 1, ..., K\}$.

4. The unitary matrix $\hat{\mathbf{U}}$ is obtained as a joint diagonaliser of the set of sample estimates $\{\hat{\mathbf{R}}_y(\tau_j) | j = 1, ..., K\}$.

5. The mixing matrix $\mathbf{A}$ is then estimated as $\hat{\mathbf{A}} = \hat{\mathbf{W}}^{\Upsilon}\hat{\mathbf{U}}$, where the superscript Υ denotes the Moore-Penrose pseudo-inverse (Penrose, 1955). The estimated source signals are then $\hat{\mathbf{Y}} = \hat{\mathbf{U}}^H\hat{\mathbf{W}}\mathbf{X}$.

In this manner, the blind source separation of temporally correlated sources is possible, based on the "joint diagonalisation" of an arbitrary set of covariance matrices. As mentioned by De Vos et al. (2011), the SOBI method offers a number of attractive features: namely that it allows (in contrast to higher order cumulant techniques such as fast ICA and ICA infomax) the separation of Gaussian sources and that due to the use of several covariance matrices, the algorithm is robust to indeterminacies.

7.6.2 Removing respiratory artefact components

In the respiratory artefact removal method outlined by De Vos et al. (2011), SOBI is used to obtain source components of the EEG signal, and (theoretically) separate EEG sources from respiratory artefact sources. The sources corresponding to respiratory artefact must

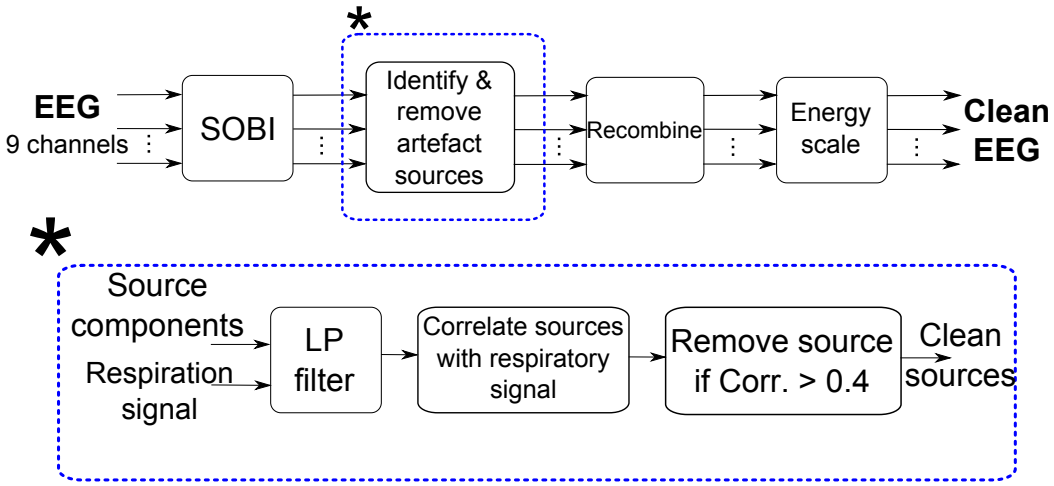


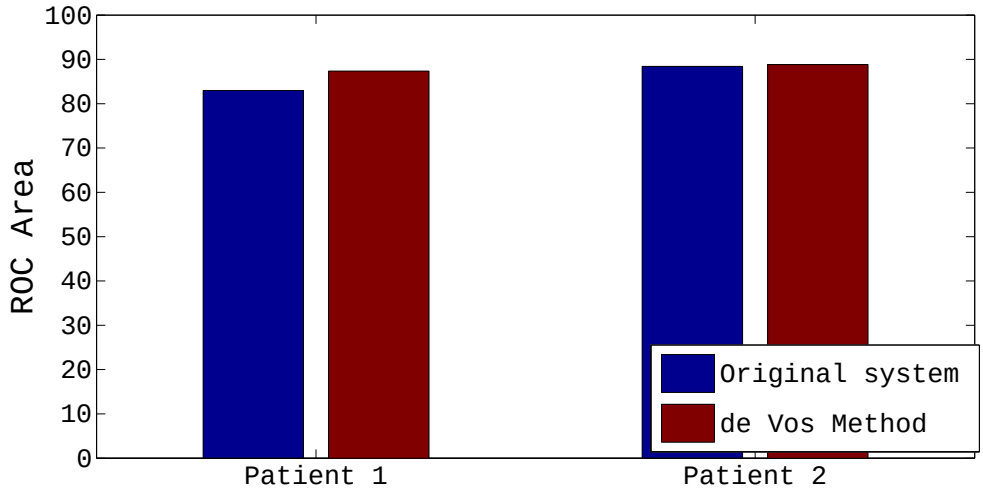
Figure 7.6: Neonatal seizure detection system with respiratory artefact removal using respiration signal included in the pre-processing stage, as outlined by De Vos et al. (2011).

then be identified; to accomplish this, each source is compared with the simultaneously recorded polygraphy signal. However, this comparison is confounded by the fact that, like ECG artefacts, respiratory artefacts can be remarkably different in morphology to those of the signal in the polygraphy measurement (Strobach et al., 1994). To overcome this problem, De Vos et al. (2011) propose an intermediary step whereby both EEG and polygraphy sources are transformed in order to enhance the similarity between them. This transformation is realised by means of a low-pass filter with frequency cut-off at 9 Hz. Identification of artefactual sources is subsequently accomplished by evaluating the correlation between the filtered EEG sources and the filtered respiration signal; if the correlation is higher than 0.4, the source is deemed artefact and is removed. Once the artefact components are correctly identified, the EEG can be reconstructed without the artefactual source(s), providing an artefact-free EEG. In this thesis, a final additional step is carried out before the (theoretically) cleaned EEG signals are presented to the neonatal seizure detection classifier. As the seizure detection classifier outlined in Section 7.2.2 includes features that depend on the energy of the EEG signal, and removing sources equates to removing energy from the EEG, this may lead to unexpected consequences when the signal is presented to the classifier. Consequently, an energy-scaling step is implemented on the "cleaned" EEG signal before feature extraction takes place. This energy scaling is accomplished by normalising the recombined EEG signal to the energy of the EEG signal before artefact removal was performed.

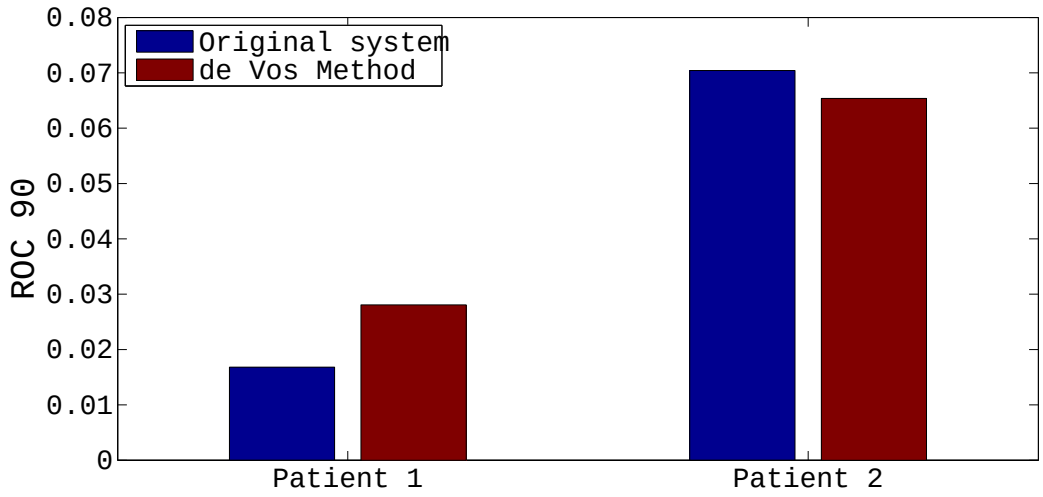
7.6.3 Algorithm performance

The De Vos method was applied to EEG recordings of two neonatal patients that were acutely affected by respiratory artefact. Comparison of the performance of the seizure detection algorithm with and without respiratory artefact removal are displayed in Figures 7.7a (ROC areas) and 7.7b (ROC_{spec90} areas). Figure 7.7a shows that the ROC area of the seizure detection classifier are increased when the respiratory artefact removal algorithm is applied in the pre-processing stage. However, for neonatal seizure detection, the area of the ROC curve with high specificities is of most interest, i.e. it is acceptable to miss some seizure epochs to accommodate reduced false detections. Figure 7.7b shows that in the area of the ROC curve with specificities above 90 %; the results for Patients 1 and 2 are contradictory. For Patient 1, the ROC_{90} almost doubles with the addition of the respiratory artefact removal, showing a relative improvement of 13.25 %. However, for Patient 2, whose ROC_{spec90} without respiratory artefact removal was more than four times that of Patient 1, disimproves slightly with the addition of the respiratory artefact removal dropping from 0.07 to 0.066. These trends are repeated in Figures 7.9a and 7.9b, which show the false detections per hour (FD/h) against good detection rate (GDR). For Patient 1, applying the respiratory artefact removal stage results in a drop in the number of false detections for an equivalent GDR; in most regions, the false detections are halved compared to the baseline neonatal system. Importantly, this trend is strongest for low and medium numbers of false detections, i.e. the areas at which the seizure detector will realistically be used. For Patient 2, FD/h increase with the application of respiratory artefact removal.

While ROC, ROC_{spec90} and FD/h vs. GDR plots are important for evaluating the expected performance of a machine learning system, each diagram shows the classification with one parameter (the threshold) as yet unspecified. In practice it is necessary to select this parameter prior to utilising the ANSeR neonatal seizure detection tool. Admittedly, in a practical implementation of a seizure detection system in the NICU, there exists the possibility of allowing the nursing staff to change this threshold value based on their interpretation of false detections (i.e. if there are too many false detections, the threshold can be increased). However, the ideal scenario would allow a threshold choice that offered sufficient performance for all patients, requiring no calibration by the nursing staff. To examine this possibility, the false detections per hour are plotted against threshold θ for Patients 1 and 2 in Figures 7.9a and 7.9b. Interestingly, for false detection rates below 1, the respiratory artefact removal preprocessing stage introduces no reduction in



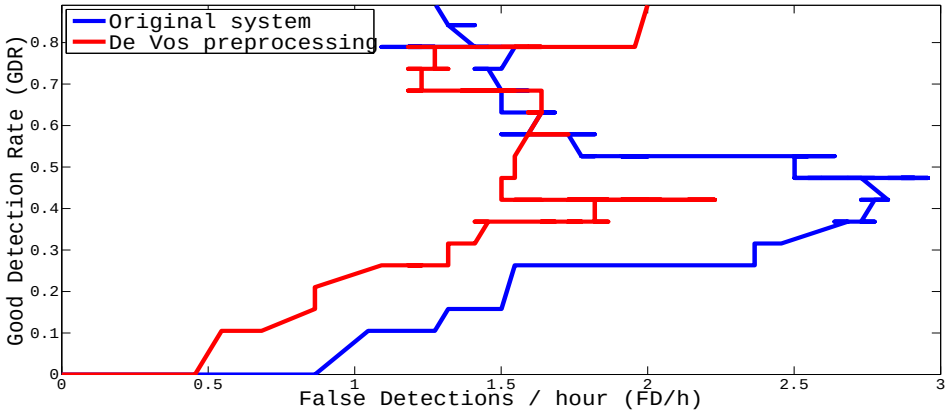
(a) ROC Area.



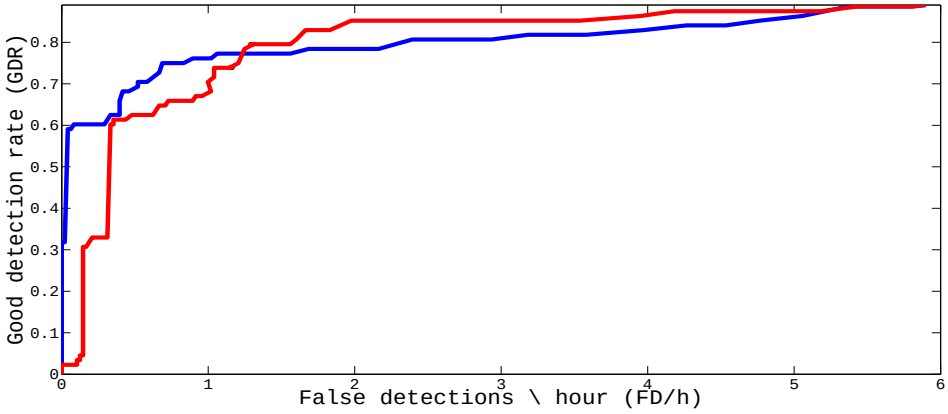
(b) ROC90 Area.

Figure 7.7: ROC and ROC_{spec90} areas corresponding to the original neonatal seizure detection performance for two neonatal patients. The blue bars show ROC and ROC_{spec90} areas without explicit artefact removal techniques. The red bars display ROC and ROC_{spec90} areas after removing respiratory artefact utilising SOBI and the respiratory trace as per De Vos et al. (2011).

performance for Patient 2, while still offering a drop in FD/h for Patient 1. Thus, while introducing the artefact removal stage reduces the GDR of Patient 2 for a given FD/h, by fixing the threshold θ, FD/h does increase; rather, the adverse effect of the artefact removal algorithm for Patient 2 can be interpreted as a drop in GDR. This may be a reasonable sacrifice for improved robustness to false detections for patient independent seizure monitoring.



(a) Patient 1.

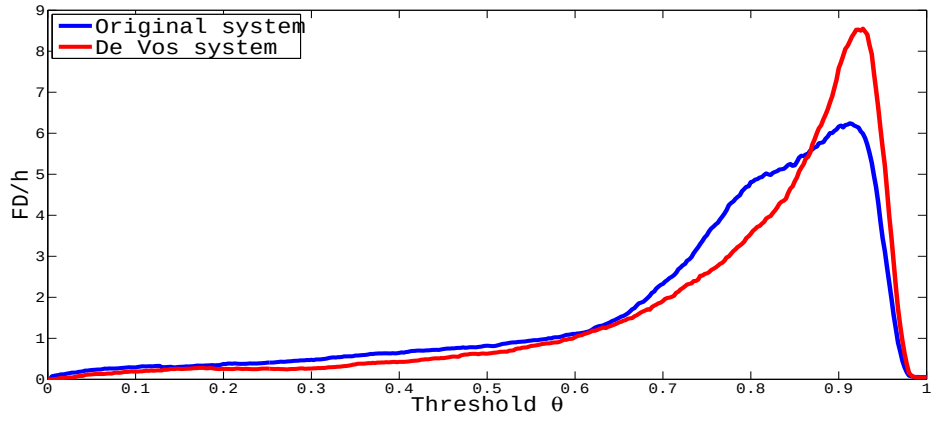


(b) Patient 2.

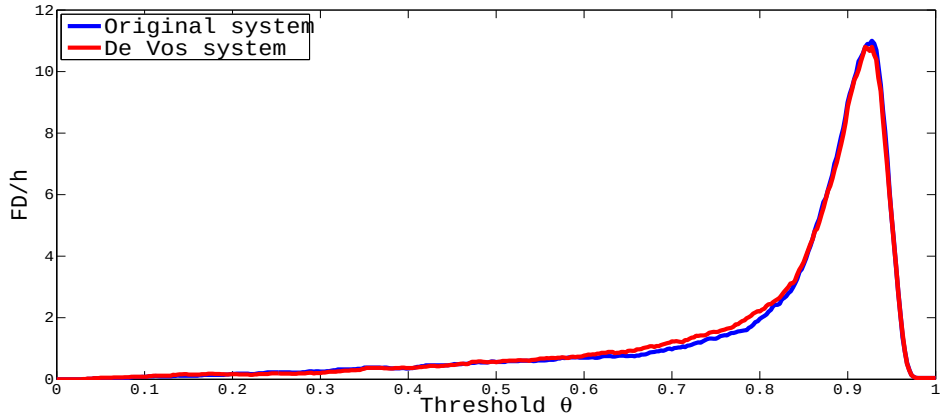
Figure 7.8: False detections per hour (FD/h) versus good detection rate (GDR) for Patients 1 and 2, for original neonatal seizure detection system (blue) and the neonatal system with respiratory artefacts removed using SOBI and the respiratory trace (red).

7.7 Respiratory artefact removal without a respiratory trace

In many neonatal intensive care units, respiratory traces are often unavailable. In NICUs where respiration is recorded, recordings are often intermittent or noisy; an issue that is strongly linked to the fact that they can cause considerable discomfort to the baby. Consequently, the ability to remove respiration artefact from the EEG without the use of a respiration trace is required if the problem of respiration artefact is to be solved for widespread neonatal monitoring. If blind source separation is to be utilised, a means of automatically identifying sources that arise due to respiration is necessary. Statistical thresholding of source components to separate between seizure and respiration artefact would be one such solution. However, initial experiments indicated that a suitable feature



(a) Patient 1.



(b) Patient 2.

Figure 7.9: False detections per hour (FD/h) versus threshold for Patients 1 and 2, for original neonatal seizure detection system (blue) and the neonatal system with respiratory artefacts removed using SOBI and the respiratory trace (red).

could not be found. More advanced classification methods based on multiple features is another alternative, but is curtailed by the need to procure and annotate source components representing seizure and respiration artefact. The solution proposed here is based on a natural physiological process known as respiratory sinus arrhythmia, whereby the heart rate is modulated by breathing. By taking advantage of this effect, the respiration signal can be derived from the ECG.

7.7.1 Electrocardiogram-derived respiration signal

The normal respiratory cycle is accompanied by changes in autonomic tone which modulate heart rate. This phenomenon, known as respiratory sinus arrhythmia (RSA), is a

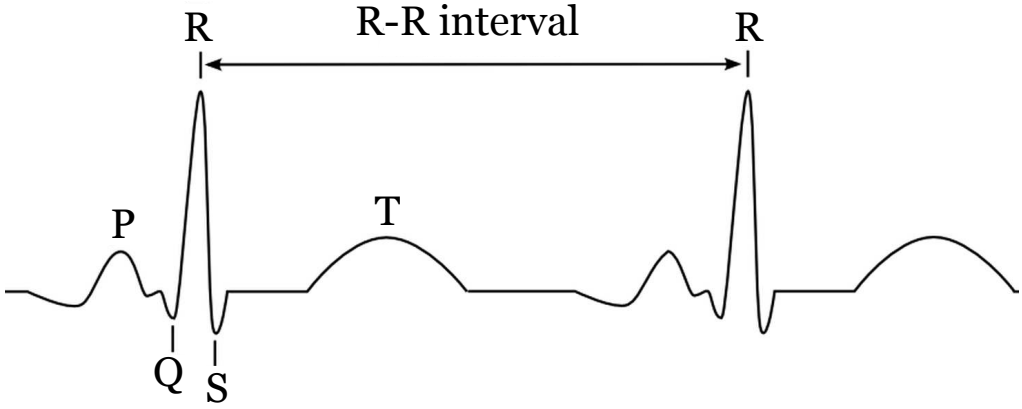


Figure 7.10: Illustration of the naming convention for ECG signal with the R-R interval displayed.

naturally occurring variation in heart-rate that occurs during a breathing cycle (Hirsch and Bishop, 1981). In this process, inhalation temporarily suppresses vagal activity, causing an immediate increase in heart rate. Exhalation then decreases heart-rate and causes vagal activity to resume. Thus, by observing beat-to-beat variations in the heart-rate, the respiratory signal can be compiled (Moody et al., 1985). The R-R interval (illustrated in Figure 7.7) is the fundamental rhythmic measure of ECG interpretation, representing the length of a ventricular cardiac cycle; it is measured between two successive R waves, and is typically used to indicate ventricular rate. The R-R interval is related to heart-rate (HR) in beats per minute (BPM) as $HR = 60/RR$, where RR is in seconds. The variation in heart-rate then can be calculated as the derivative of the R-R interval. This method of obtaining an electrocardiogram-derived respiration (EDR) signal is perhaps the most straightforward to implement; yet as outlined by Boyle et al. (2009), it offers similar levels of performance to more complex techniques based on characteristics of beat morphology such as area under the QRS complex (Moody et al., 1985), amplitude of the R-wave (Khaled and Farges, 1992) and amplitude of the R- and S- waves (Mason and Tarassenko, 2001).

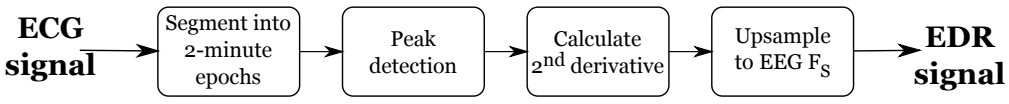


Figure 7.11: Algorithm for obtaining electrocardiogram-derived respiration signal as outlined by Boyle et al. (2009).

The electrocardiogram-derived respiration (EDR) algorithm used here thus requires two main steps: detection of the QRS complexes and calculation of the variation in the R-

R intervals between these QRS complexes. QRS detection is a well-established research area with popular algorithms based upon the derivative of the ECG signal (Pan and Tompkins, 1985), wavelets (Afonso et al., 1999), neural networks (Barro et al., 1998) amongst others. In this work, the wavelet-based QRS detection algorithm by Afonso et al. (1999) was utilised using the Biosig toolbox in Matlab (Vidaurre et al., 2011). QRS detection is thus performed by decomposing the ECG into non-overlapping frequency bands and extracting a set of features based on the power within these ranges; beat classification is then performed using a set of thresholding routines and heuristics. Before the electrocardiogram-derived respiration (EDR) algorithm is implemented, the ECG signal is segmented into non-overlapping windows of 2 minutes duration. The EDR algorithm is then carried out on each window as follows:

1. Perform automated QRS peak detection on the ECG signal, as outlined in Afonso et al. (1999).
2. Obtain the variation of the R-R intervals, i.e. the second derivative of the peak positions.
3. Upsample the EDR signal to that of the EEG signal (250 Hz), so that correlation can be performed between the EDR and the each of the source components.

Figure 7.12 displays a section of respiration trace and its equivalent EDR signal for Patient 1. Note that the same number of peaks are shown for both respiration and EDR signals, with a slight phase change.

7.7.2 Respiratory artefact removal with the EDR signal

The electrocardiogram derived respiratory signal offers a potential solution to the problem of automatically identifying source components associated with respiratory artefact. As a surrogate respiration signal, it is proposed here to substitute the EDR signal for the respiration trace in the De Vos method outlined in Section 7.6. The neonatal seizure detection algorithm with the EDR signal would thus look like that outlined in Figure 7.13. To account for the phase shift between respiration and EDR signals and consequently fall in correlation between the EDR and respiration artefact on the EEG, the correlation threshold θ was reduced from 0.4 to 0.2. This single parameter was chosen heuristically

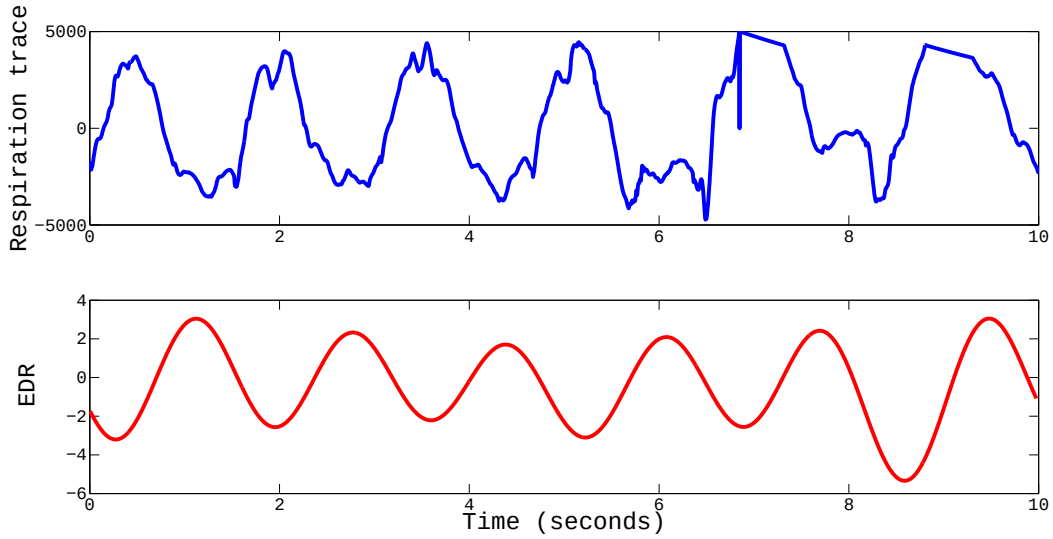


Figure 7.12: Respiratory trace and equivalent EDR signal for 10 seconds of data for Patient 1.

(on a patient from a separate dataset); a thorough parameter search may yield an optimal choice.

The seizure detection performance of the respiratory artefact removal using the EDR signal is compared to that of the baseline seizure detection system as well as that of artefact removal using a respiratory trace in Figures 7.14a and 7.14b. For both ROC area and ROC_{spec90} for both patients, performance of the artefact removal with the EDR signal approaches that of the artefact removal with the respiratory trace. Consequently, ROC area and ROC_{spec90} area are improved for Patient 1 when compared to the baseline neonatal seizure detection system without respiration artefact removal. For Patient 2, the ROC area with and without respiratory artefact removal are equivalent. As was the case for the artefact removal using the respiration trace, ROC_{spec90} area decreases when artefact removal using the EDR signal is implemented on Patient 2.

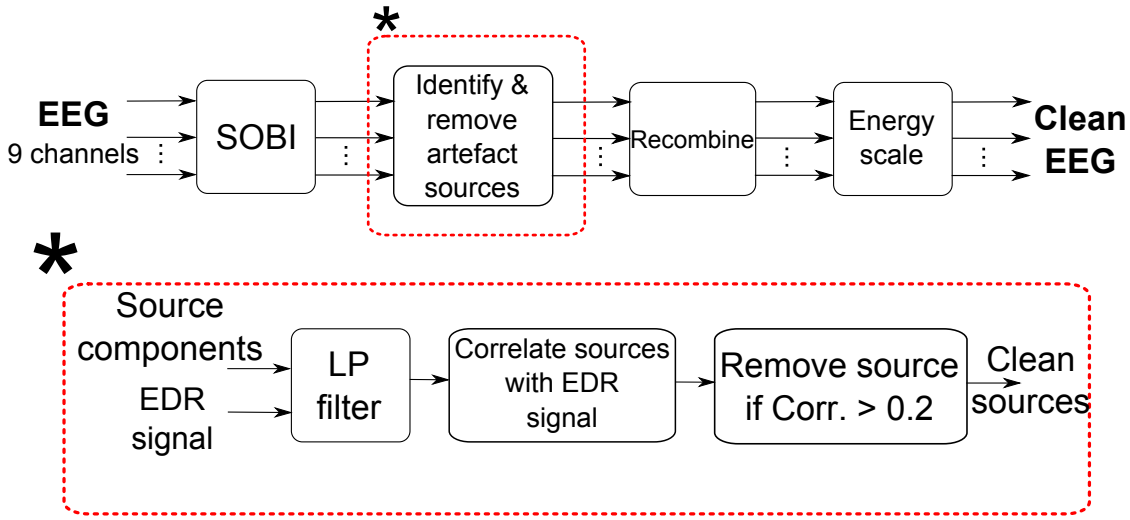
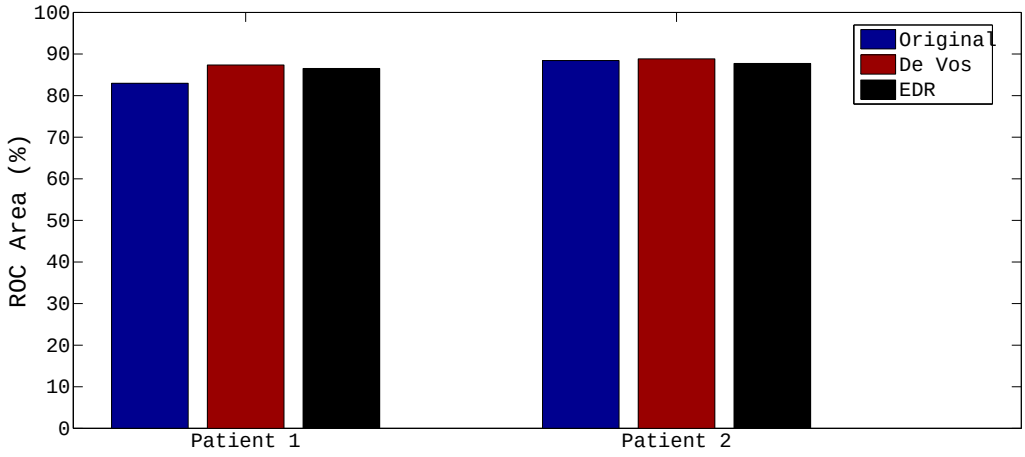


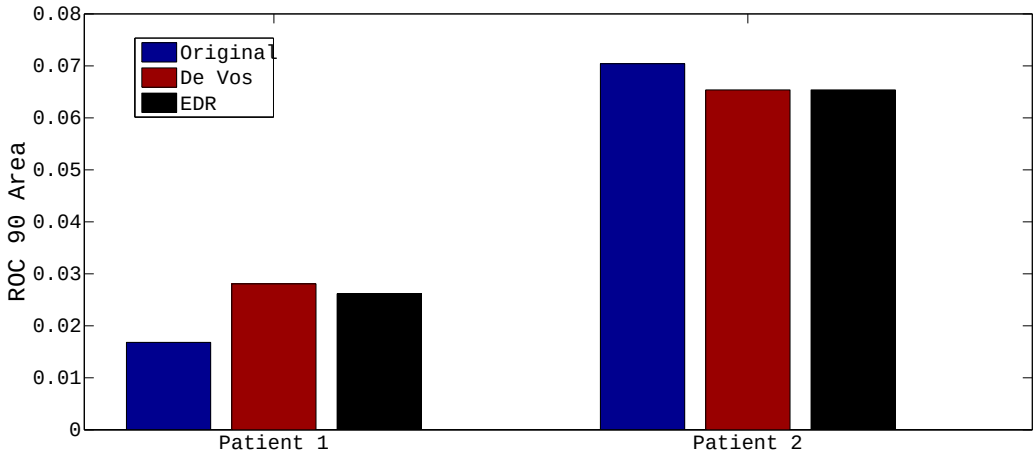
Figure 7.13: Neonatal seizure detection system with respiratory artefact removal using EDR signal included in the pre-processing stage.

The ROC performance of the respiration artefact removal algorithm using the EDR trace is evaluated for two additional patients (Patients 3 and 4) in Figures 7.15a (ROC area) and 7.15b (ROC_{spec90} area). In these patients, no respiration trace was recorded. It is evident in the barcharts, that the seizure detection performance as evaluated by ROC area improves with the addition of respiration artefact removal using SOBI and the EDR signal. In Figure 7.15b, the ROC_{spec90} area before and after artefact removal using the EDR signal is illustrated. For Patient 3, a relative improvement of 6 % is observed with the artefact removal stage. However, Patient 4 does not register any significant improvement in seizure detection performance in this area of the ROC curve with respiratory artefact removal.

False detections are examined in Figures 7.16a, 7.16b, 7.16c and 7.16d. Performance of Patients 1 and 2 is similar to those using the De Vos algorithm, albeit with a reduced improvement here for Patient 1 and a reduced disimprovement seen here for Patient 2. Patient 3 improves with addition of respiration artefact removal using the EDR signal, with FD/h falling or remaining constant for all GDR values below 0.7. The same trend is evident for Patient 4, with the most notable improvement seen for FD/h below 0.5. In conclusion, the EDR method is a suitable respiration artefact removal method for use in the ANSeR system where a respiration trace is unavailable.



(a) ROC area.

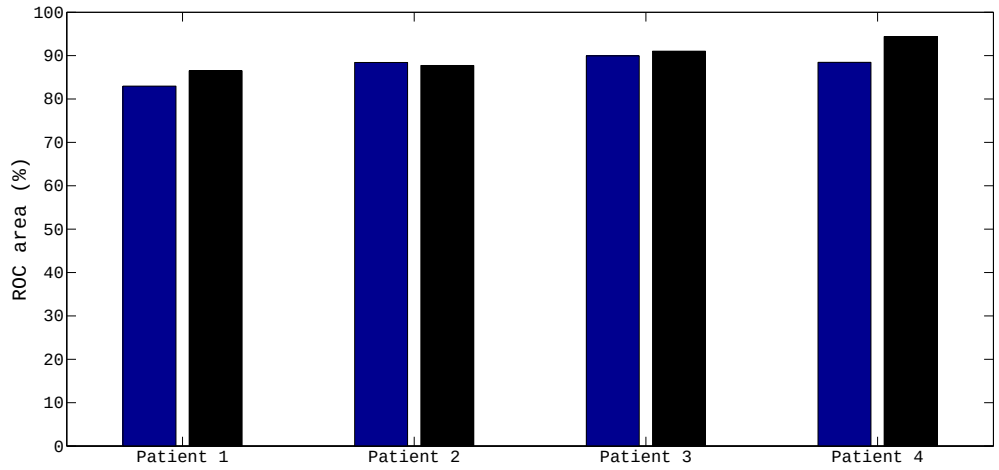


(b) ROC 90 area.

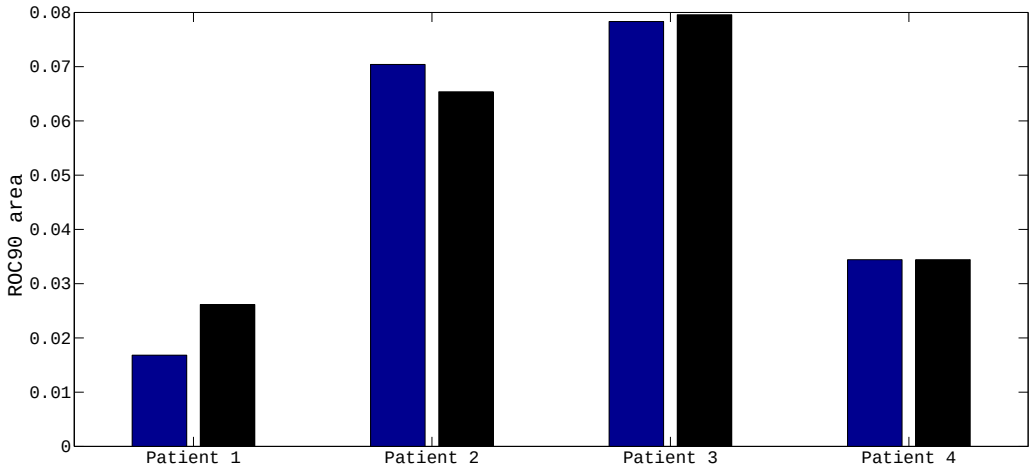
Figure 7.14: ROC and ROC_{spec90} areas showing original neonatal seizure detection system (blue), and the system with respiratory artefact removed using the respiratory trace (red) and EDR signal (black).

7.7.3 Artefacts on additional physiological signals

The respiration trace and ECG signal used to identify respiration artefact on the EEG are also prone to artefacts themselves. Consequently, artefact detection methods which rely on additional physiological signals are susceptible to reduced performance by the presence of artefacts on these auxiliary physiological signals. If there are significant amounts of artefacts on the additional physiological signals, the artefact removal algorithm may compromise the neurological event detection system and negatively affect the classification performance. In this experiment, this was found to be case for sections of the recordings.



(a) ROC area.



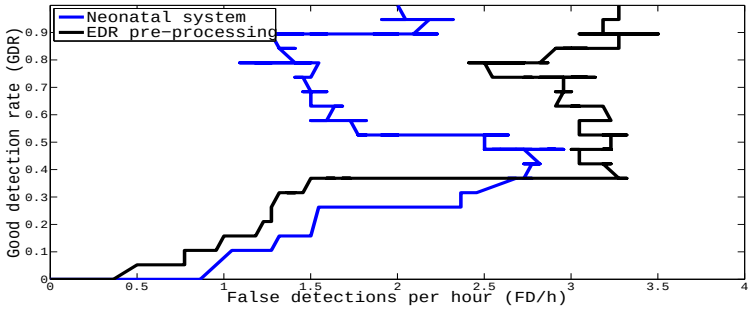
(b) ROC 90 area.

Figure 7.15: ROC and ROC$_{spec90}$ areas showing original neonatal seizure detection system (blue), and the system with respiratory artefact removed using the electrocardiogram derived respiratory (EDR) signal (black). ROC and ROC$_{spec90}$ areas are reported for four neonatal patients that underwent seizure and whose EEG was contaminated by periods of significant respiratory artefact.

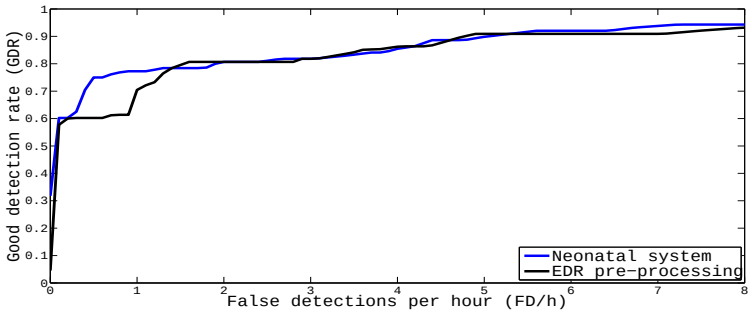
Figure 7.17 displays an example of artefact on the respiratory signal. In this work, steps were taken to ensure that artefact removal was only performed if artefacts did not appear on the additional physiological signals. For both the De Vos and EDR methods, artefact removal was utilised only if the respiration trace did not contain extended periods of low amplitude signals (as per the number of inactive samples described in Chapter 3) or extended periods of high amplitude artefact (as measured by samples more than 2 standard deviations above the mean).

7.8 Summary

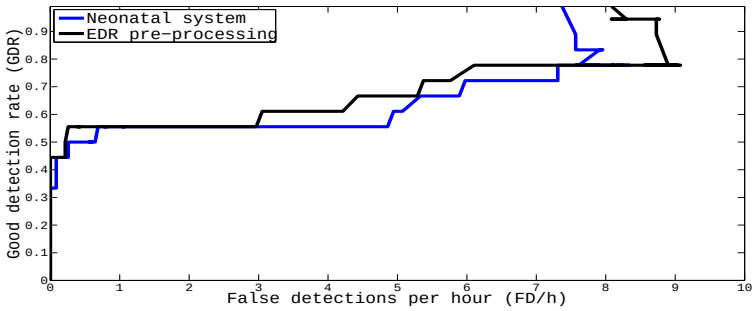
In this chapter, two methods to remove respiratory artefact from neonatal EEG using blind source separation were investigated. The De Vos method of incorporating additional physiological signals in the form a respiratory trace was successful at removing respiratory artefact from the EEG signal, and subsequently improving the performance of the ANSeR neonatal seizure detection system. A novel approach to respiratory artefact removal was also introduced which made use of a respiration signal derived from the ECG. In doing so, the need for a respiratory trace is circumvented, allowing respiratory artefact removal in EEG recordings where respiration is not monitored. As the majority of NICUs do not monitor respiration, the algorithm developed in this chapter represents the only available method of respiratory artefact removal for many neonatal monitoring units. The improvement in performance was most pronounced for patients whose EEG contained considerable numbers of false detections due to artefact. For patients with already (comparably) low numbers of false detections, the seizure detection performance was unaffected or in the case of one patient, slightly worsened.



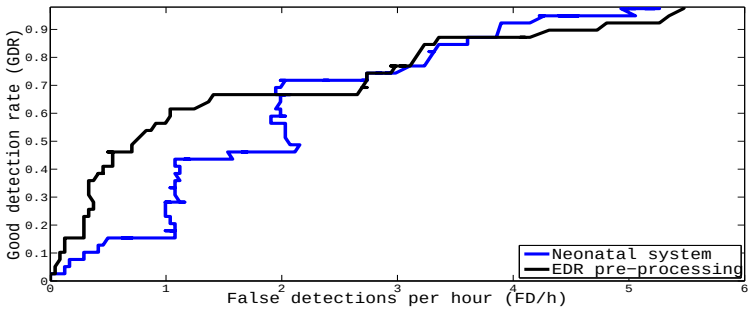
(a) Patient 1.



(b) Patient 2.



(c) Patient 3.



(d) Patient 4.

Figure 7.16: False detections per hour (FD/h) versus good detection rate (GDR) showing original neonatal seizure detection system (blue), and the system with respiratory artefact removed using the electrocardiogram derived respiratory (EDR) signal (black). Results are reported for four neonatal patients that underwent seizure and whose EEG was contaminated by periods of significant respiratory artefact.

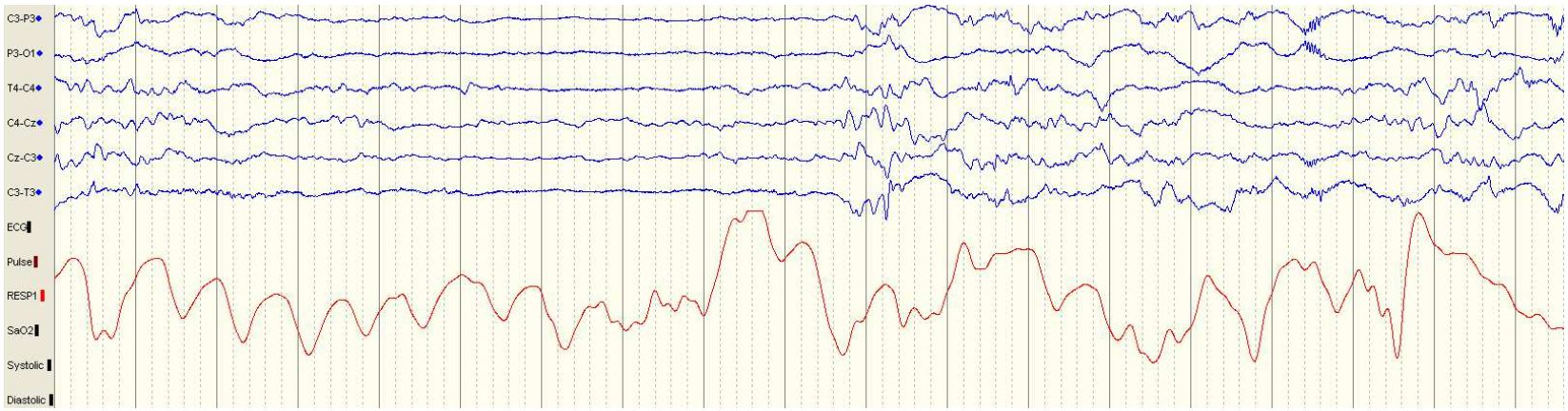


Figure 7.17: Example of motion artefact on the respiratory trace (red) of a neonatal recording at University College London Hospitals. The respiratory trace is relatively artefact-free in the left half of the image; motion artefact causes high amplitude deviations in the right half of the image in both the respiratory trace (red) and EEG signal (blue).

Chapter 8

Conclusions and future work

8.1 Conclusions

The primary aim of this thesis has been to use biomedical signal processing and machine learning techniques to mitigate the deleterious effects of EEG artefacts in real-world environments. Specifically, these techniques have been deployed to detect and remove EEG artefacts so as to allow the operation of automated neurological event detection systems in these real-word settings; in ambulatory domains, in neurology departments and in intensive care units. The approach taken with this research has been to focus on those artefacts that are most problematic, i.e. those artefacts that prevent an EEG technology from deployment. In this regard, head-movement artefact proved to be the most troublesome source of contamination in ambulatory environments. Similarly, detailed analysis of the state-of-the-art epileptiform activity detection and neonatal seizure detection systems showed that movement and ocular artefact caused the most false detections in routine epilepsy monitoring, and respiration artefact led to the most false alarms by the ANSeR seizure detection system. Accordingly, attempts to alleviate the difficulties introduced by these artefacts makes up the contributions of this thesis.

The work in this thesis has approached the problem of artefacts primarily from two directions: (i) utilising supervised machine learning methods to accurately detect artefacts and incorporating this procedure into the decision making of the automated neurological event detection system and (ii) utilising blind source separation techniques allied with information from additional physiological signals to remove the contribution of artefact.

In Chapter 2, a detailed review of the existing literature on artefact processing revealed a number of shortcomings in their treatment in ambulatory EEG. Notably, ambulatory artefact detection techniques were either non-existent or were evaluated using simulated artefact data. In automated neurological event detection systems, the existing literature fell short in a number of ways: the effect that artefact processing methods had on the EEG diagnostic system was not quantified, and the methods were either outdated (simple amplitude thresholding) or implemented with systems that have long-since ceased being state-of-the-art. In neonatal seizure detection, the respiration artefact removal techniques relied upon the use of a respiration trace, an entity uncommon in many NICUs. Chapter 2 went on to detail the EEG data used in this thesis; to the author's knowledge this is the most comprehensive collection of representative EEG data in the literature with which to evaluate artefact processing techniques.

Chapter 3 outlined the problem of contamination by head-movement artefact in ambulatory EEG. A statistical machine learning approach was taken to the problem of identifying EEG sections that are contaminated by electrical activity arising from non-cerebral sources. Inspection of a comprehensive feature set (comprising time domain, frequency domain and information theoretic features) indicated their usefulness in the generalised head-movement artefact detection task. The approach for head-movement artefact detection developed here was the first work to deal with artefacts in an ambulatory environment and in doing so, to class all movement related artefacts together. This concept of generalised head-movement artefact detection was initially validated using a reduced feature set and a linear discriminant classifier; classification accuracies of between 65 % and 70% were recorded for each of the window lengths investigated with a window length of 1.5 seconds showing the best performance. With the feasibility of detecting head-movement artefacts with a single classifier confirmed, classification using a support vector machines classifier was performed utilising the full feature set. Mean ROC areas of between 75 % and 83 % were observed, this time on the more difficult task of participant independent artefact detection. A window length of 0.75 seconds resulted in the highest mean ROC area of 83 % ; however, as for the linear discriminant classifier, given the variation in SVM classification performance across participants, a definitively best window length was not chosen. The SVM classifier trained and tested in Chapter 3 is suitable for artefact annotation of EEG recorded in an ambulatory environment, and to the author's knowledge is the first of its kind. Additionally, this artefact detector could be incorporated into an automated neurological event detection system at either the pre-processing (via artefact rejection) or post-processing stage (via classifier fusion).

The work in Chapters 4 and 5 took advantage of the miniaturisation of gyroscopes, and used this additional physical signal to detect artefacts in the EEG; this work was the first to use gyroscopes for this purpose. Results from the mutual information calculations and linear discriminant classifiers indicated that the gyroscope feature set was useful in discriminating between resultant EEG artefact and non-artefact classes. LDC classification accuracies of between 65 % and 68 % for the 4- and 5- feature LDCs validated the use of gyroscope features to detect non-cerebral, movement-induced activity in the EEG. As was observed for the EEG classifier, utilising an SVM classifier on a comprehensive feature set for the gyroscope classification task lead to good separation between gyroscope signals during normal EEG and those during head-movement artefact, providing a strong argument for including gyroscopes in an EEG artefact detection system. Indeed, the gyroscope classifiers performed at similar levels to the EEG classifiers (mean ROC areas between 80 % and 84 %) and even outperformed them for some window lengths. Once more, however, a single best window length was not chosen, due to the variability in mean ROC areas across participants for each window length investigated. Chapter 5, investigated methods of combining the EEG and gyroscope signals in the detection of artefacts arising from head movements. In doing so, the performance of the artefact detection system was improved and a framework was developed whereby additional physiological signals can be incorporated into the artefact detection task. It was found that the fusion of signals at both feature and classifier levels improved detection of head-movement artefact in ambulatory EEG when compared to using either EEG or gyroscopes alone. This result was observed for feature fusion as well as for each of the score-level and decision-level fixed-rule combinations that were investigated. Amongst these methods, it was found that feature fusion and the score-level, sum rule offered the best classification performance by increasing the ROC areas and reducing inter-participant variability. Improvements in mean ROC area for a window length of one second was 8 % for the feature fusion and score fusion sum-rule combination methods when compared to the EEG classifier alone. These results confirm the complimentary nature of information carried by these different modalities. Thus, in order to most effectively detect head-movement artefacts in EEG, a classifier that combines EEG and gyroscope features at either the feature or score level is recommended.

Chapter 6 focused on improving the classification performance of a state-of-the-art epileptiform activity detection algorithm. To achieve this, SVM classifiers that discriminate between epileptiform activity and the artefact types that lead to the most false detections (namely, ocular and movement artefacts) were employed in a post-processing classifier

fusion stage. Using both epoch and event-based metrics, it was clear that the use of artefact-specific SVM classifiers in EEG sections of suspected epileptiform activity improved the performance of the automated epileptiform detection system. The FDR metric verified that a considerable reduction in the number of false detections was achieved whilst continuing to accurately detect epileptic seizure and short-duration, interictal epileptiform events; median false detection rate fell by 83.27 % at a constant good detection rate of 0.9375. The most profound improvement was found by incorporating head-movement and ocular artefact detection classifiers together and combining them with the baseline epileptic seizure and epileptiform activity detection classifier. A further benefit of utilising the artefact SVMs was that the classification performance gains were seen across patients, adding a level of robustness to the epileptiform detection algorithm that is necessary in a real-world setting. In light of these results then, the use of score level classifier fusion for epileptiform activity detection is recommended as it offers superior performance to that of the decision level classifier fusion, reducing median false detection rate from 0.3 to 0.05.

Chapter 7 outlined efforts to deal with false detections due to respiration artefact in the state-of-the-art ANSeR neonatal seizure detection system developed at U.C.C. To achieve this, blind source separation techniques using SOBI and correlation of the source signals with the respiration trace, of the form used in De Vos et al. (2011), were implemented. To circumvent the issue of lack of widespread availability of respiration signal in NICUs, a novel respiration algorithm was developed utilising an electrocardiogram-derived respiration signal. This artefact removal algorithm improved classifier performance in patients where respiration artefacts lead to significant false alarms. The improvement in performance was most pronounced for patients whose EEG contained considerable numbers of false detections due to artefact.

8.2 Future research directions

Research is always incomplete; there will always be more to do. Why then stop here? As stated above, the aim at the outset of this work was to further advance EEG artefact processing to a point where substantial progress was made in bringing automated EEG medical technologies to bear on real-world applications. In this respect, artefact processing in three broad areas were advanced. In the first body of work, namely artefact detection in ambulatory EEG, it is felt that the bulk of improvements from utilising multimodal, supervised machine learning have been gleaned from the available data. Additional advances

will be increasingly incremental, and in reality will require the collection of new datasets. This is discussed in more detail in Sections 8.2.1 and 8.2.2. In the second body of work, i.e. artefact classifier fusion in the context of epileptiform activity detection, significant improvements in performance have been realised. Further improvements are expected, but as explained in Section 8.2.3 this will likely require additional well-annotated data or investigating fundamentally alternative approaches. Finally, in the third body of work, that of respiratory artefact removal in a neonatal seizure detection system, the improvements made, appear sufficient in bringing classifier performance on those patients who previously performed poorly to a level, where false detections are acceptably small. Details of suggested future work are outlined in Section 8.2.4.

At the same time, the research outlined in this thesis raises further questions. The following paragraphs outline some areas that may offer interesting directions for future work.

8.2.1 EEG head-movement artefact detection

In all classification tasks there remains the possibility, albeit unlikely, of discovering one "super-feature"; a feature which provides almost perfect discrimination between the class labels in question. Perhaps more realistically, there exists ample scope for the discovery of features in both the EEG and gyroscope signals that would provide incremental improvement in classification performance. Recently, Temko et al. (2011a) verified the applicability of automated speech recognition (ASR) features in neonatal seizure detection. An investigation of the usefulness of these features in the artefact detection task would be interesting. For the gyroscope signals, the activity detection literature offers potential for seeking additional, discriminative gyroscope features.

8.2.2 Multimodal head-movement artefact detection

The score fusion methods investigated in Chapter 5 used fixed combining rules. However, there are several reasons why the outputs of the individual EEG and gyroscope classifiers may not be optimally scaled with respect to each other (Duin, 2002). There exist a number of potential methods by which the EEG and gyroscope classifiers may be combined to provide further improvement in the head-movement artefact detection task by accounting for this fact. These methods involve a further evaluation data set which is used to vary several parameters in the classifier combination process. Future work would include the

expansion of the artefact database used in this experiment, so that weighted and trainable classifier combination can be investigated. Finally, the head-movement artefact detection systems outlined in Chapter 5 are intended for use in ambulatory EEG for medical and BCI applications. The manner in which artefact detection is incorporated into those systems should also be explored. To this end, future work could include the collection of data comprising labelled artefact and medical/ BCI events, and should explore ways in which head-movement artefact detection can improve the medical/ BCI classification task. This was attempted to a degree in the work carried out in Chapter 6; however, ambulatory EEG and gyroscope data collected and annotated from epileptic patients would offer the most realistic view of how artefact processing can be incorporated into continuous ambulatory EEG monitoring of patients with suspected epilepsy.

The fusion methods outlined in Chapter 5 introduced the concept of combining EEG and gyroscope signals to improve the detection of head-movement artefacts in EEG. There exist a number of additional physiological signals which are promising in the detection of other EEG artefacts. EOG, EMG and respiration signals could all be incorporated into the artefact detection system by means of score fusion of individual classifiers as outlined in this thesis.

In an ambulatory EEG system, minimizing the computational burden of feature extraction and classification is likely to be a key, practical issue. This becomes even more critical when we consider that artefact detection and removal is likely to be a secondary function in any automated EEG neurological event detection system. It follows, that in a case where computational load is limited, artefact detection will surely be amongst the first components to be scaled back. Exploration of the effect of reducing computational burden could take place at the feature level (by means of recursive feature elimination (RFE)) or at the classifier level via support vector reduction.

8.2.3 Artefact processing in automated epileptiform activity detection systems

The work outlined in Chapter 6 aims to combat the detrimental effect of ocular and movement artefacts in the epileptiform activity detection task. The proposed method of classifier fusion of epileptiform and artefact SVMs proved extremely effective at reducing the number of false positive detections while maintaining a high good detection rate. However, the underlying source of false detections are not limited to ocular and movement

artefacts. Other artefact types (such as sweat, ECG and respiration) can also be blamed for false positive detections of epileptiform activity. Expanding the dataset to include labelled examples of additional artefact types would allow the training of an array of artefact-specific classifiers and could further reduce the number of false detections in the epileptiform activity detection task. Such a system would then conceivably possess a range of classifiers to discriminate between epileptiform activity and each of a number of artefact types. In doing so, it is proposed that the number of false detections can then be reduced to levels which would allow an automated epileptiform detection system to be deployed in a clinical environment. Alternatively, the approach taken in Chapters 3, 4 and 5, where artefacts are detected for all epochs, not just those epochs containing the suspected neurological event in question, could be investigated here, and all artefacts could be included in one artefact class. Similarly, a comparison of artefact processing at the post-processing stage with artefact rejection at the pre-processing stage (as discussed in Chapter 3 would be a useful addition to the EEG artefact literature.

The score fusion method outlined in Chapter 6 is a fixed-rule, classifier combination method. Trainable classifier combination, whereby the posterior probabilities from the base classifiers are used as input features to a general classifier used for combining, have been shown to outperform fixed-rule combination for many applications. The use of these trainable combining classifiers typically requires larger datasets than the equivalent fixed-rule combiner. The expansion of the epileptiform database and the subsequent investigation of trainable classifier combinations provides another interesting direction for future work. It is hypothesised that employing a trainable combining classifier would lead to a further reduction in false positive detections and could be included with the aforementioned range of artefact-specific classifiers.

Finally, the definition of artefacts assumed in this thesis is that of any electrical activity appearing on the EEG that arises from extra-cerebral sources. However, an alternative definition of artefact, as "any electrical signal appearing on the EEG which interferes with the classification task in question", may pose some interesting questions. By re-framing the artefact definition in this manner, cerebral activity such as alpha and delta waves fall under the umbrella term of EEG artefact. As alpha and delta brain activity have been documented as causes of false positive detections in the epileptiform activity detection task (Gotman and Gloor, 1976), classifiers that are tailored to discriminate between epileptiform activity and alpha/delta activity could be included in the artefact processing system and potentially lead to further reductions in false detection rate.

An alternative approach to performing classifier fusion of multiple discriminative classifiers at the post-processing stage would be to use multi-class generative classifiers such as Gaussian mixture models (GMM). This family of classifier models are widely used in speech processing for multi-class classification. Within speech processing, GMMs have proven particularly useful in removing various forms of non-speech sounds, i.e. they are effective for dealing with speech processing artefacts.

8.2.4 Respiratory artefact removal in automated neonatal seizure detection systems

The work carried out in Chapter 7 represents the initial progress achieved in removing respiration artefact from neonatal EEG using an EDR signal. There exists a number of ways in which the developed system could be optimised. Firstly, the correlation threshold coefficient was chosen heuristically by examining a short portion of data from a patient external to the test data set. By enlarging the data set, nested cross validation could be used to select an optimal correlation threshold.

The EDR algorithm chosen was that which appeared straightforward to implement and which offered a clean signal. There is however, considerable scope for iterative improvement by utilising more advanced EDR methods such as kernel PCA. Similarly, more sophisticated algorithms for beat-to-beat detection of the ECG signal may offer improvement in the quality of the EDR signal and subsequent improved in respiration artefact removal. Similarly, incorporation of artefact processing techniques on the ECG signal would also be beneficial to the performance of the respiration artefact removal.

8.3 Final remarks

This thesis is the culmination of years of research with the Biomedical Signal Processing Group at UCC and multidisciplinary collaborations with medical staff at CUH and UCLH, with whom the author has been grateful to work. Similarly, the work builds upon years of research on automated EEG diagnostic systems for medical applications at UCC. Put simply, without these multidisciplinary collaborations and without the knowledge and expertise that has preceded this research, this thesis would not have been possible. In this context, it is hoped that this work brings the research one step closer to deployment in

real clinical environments and thus improve the standard of care for those with suspected neurological illness. Additionally, it is hoped that the research outlined in this thesis will provide a useful platform for further advancement in the processing of artefacts in medical EEG diagnostic systems, for epilepsy and neonatal seizure detection as well as in other medical applications such as sleep analysis and Parkinson's detection. Finally, by effectively dealing with artefacts, the use of EEG technologies in ambulatory environments will allow the exponential growth for EEG monitoring in non-clinical environments that may bring hitherto unrecognised benefits to the population. It is hoped that the work detailed in this thesis will go some way in making this a reality.

Acknowledgements

Thank you to the following individuals who without their contributions and support this thesis would not have been written:

I would like to express my sincerest gratitude to Dr. Liam Marnane who offered me the opportunity to undertake this Ph.D., for encouraging me to travel and for giving me the freedom to pursue various research avenues. Similarly, I would like to thank Dr. Gordon Lightbody and Dr. Geraldine Boylan for welcoming me to the Neonatal Brain Research Group and allowing me to continue my research there.

I would like to thank SFI for funding this research, through the EEDSP and NEOPRISM projects. I firmly believe that by continuing to fund applied medical research that significant improvements can be made in the quality of life of patients both here in Ireland and abroad.

I would like to thank all the staff and students of the Department of Electrical and Electronic Engineering for creating a working environment that was enjoyable and challenging for the past eight years. I would also like to thank the administration staff; in particular, Rita Sarteschi, Geraldine Mangan and Mary Costello as well as Ralph O'Flaherty, without whom efficient work in the department would have been impossible. I would also especially like to thank Dr. Michael Egan for his advice and wisdom over the years, and in encouraging me to choose to pursue the Ph.D.

I would like to thank all the postgraduate students in the department, whose constant pessimism and frustration with Ph.D. life offered much-needed consolation. In particular, the lads from my year for tea, pints and trips away. I would especially like to thank Daniel Kelleher who collaborated on much of the research within this thesis, and who spent many hours helping with debugging Matlab code.

This thesis is very much a data-driven approach to artefact detection and removal. In particular I would like to thank everyone involved in collecting and annotating the clinical datasets: notably, Dr. Geraldine Boylan, Dr. Daniel Costello, Dr. Brian McNamara, Derek Nash, Daniel Kelleher, Robert Goulding, Dr. Nathan Stevenson and Sean Mathieson. I would also like to extend my thanks to everyone who helped in the collection of the ambulatory artefact data; notably Dr. William Hutch, Dr. Catherine Pettigrew and Lara Alegre, as well as all the willing human guinea-pigs who were instructed to perform

ridiculous movements in the hope of capturing accurate artefact data. I would also like thank Geoff Mackellar at Emotiv for helping with the headset and promoting my research.

I would like to thank the postdocs for answering my many questions and finding flaws in my arguments; Dr. Stephen Faul and Dr. Andrey Temko, this research is much stronger as a result of your input. I would like to thank the Ph.D. students in the biomedical signal processing group that went before me for building the platform upon which much of this research is based, and who helped answer questions and provide code: Dr. Eoin Thomas and Dr. Orla Doyle.

I would especially like to thank the Machine Learning Group at TU Berlin for inviting me to visit their group: Dr. Paul von Bünau, Dr. Frank Meinecke, Dr. Franz Király, and Dr. Klaus-Robert Müller.

Finally, I would like to thank my family, especially my parents, for offering support through the last four and a half years in particular, and encouraging me to move home to finish up the research. Hopefully, you'll never hear the words "artefact detection" ever again.

Statement of originality

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institute of higher learning, except where due acknowledgement is made in the text.

Simon O'Regan

April, 2013

List of Tables

2.1 Artefact Generation Protocol. 46

2.2 Data characteristics for each patient record used in the development of the epileptiform activity detection system. 47

2.3 Data characteristics for each neonatal patient used in the development of the neonatal artefact removal algorithms. 48

3.1 List of EEG features for head-movement artefact detection. 60

3.2 Top performing EEG features for the artefact detection task as ranked by MIEF for window lengths of (a) 1 second and (b) 1.5 seconds 80

4.1 List of gyroscope features for head-movement artefact detection. 99

4.2 Best-performing features using MIEF for window lengths 0.25 to 4 seconds. 105

6.1 List of features for epileptiform activity detection. 138

7.1 Breakdown of false detections by the ANSeR seizure detection algorithm due to respiratory artefact as reported in Mathieson (2012). 153

List of Figures

2.1 EEG electrodes placed according to the 10-20 system. 8

2.2 Examples of normal background EEG for an adult patient (a) and a neonatal patient (b). (a) Normal background adult EEG recorded using a Viasys Nicolet EEG machine in the Department of Neurology at Cork University Hospital. The displayed EEG is sampled at a frequency of 250 Hz. (b) Normal background neonatal EEG recorded using a NicOne EEG machine in the neonatal intensive care unit at Cork University Hospital. The EEG is sampled frequency of 256 Hz. 9

2.3 EEG sample showing a number of instances of eye blinks and lateral eye movements. The EEG was recorded using a Viasys Nicolet EEG machine in the Department of Neurology at Cork University Hospital. The displayed EEG is sampled at a frequency of 250 Hz. 11

2.4 EEG sample showing artefacts arising from head movements.The EEG was recorded using a Viasys Nicolet EEG machine in the Department of Neurology at Cork University Hospital. The displayed EEG is sampled at a frequency of 250 Hz. 12

2.5 Loose electrode on channel T4-C4, resulting in high frequency noise and pulse artefact. The ECG trace (red) shows the QRS complex preceding the ECG by 200-300 ms. 13

2.6 Pulse artefact on neonatal EEG recording; visible on channels C4-Cz and Cz-C3 (red) manifesting as a slow wave. The ECG trace (black) shows the QRS complex preceding the pulse artefact by 200-300 ms. 14

2.7 Respiration artefact, manifesting as slow waves, on neonatal EEG recording; electrode C4 records the artefact which is then visible on channels F4-C4 and T4-C4, and to a lesser extent on channels C4-P4 and C4-Cz. 16

2.8 Respiration artefact on neonatal EEG recording; visible on channels F4-C4, C4-O2, T4-C4 and C4-Cz, manifesting as slow and sharp waves. 17

2.9 Electrodermal artefact on neonatal EEG recording, most recognisable by the slow, baseline drifts on channels F4-C4, F3-C3, and C3-T3. 18

2.10 A basic representation of learning from data. 21

2.11 EEG showing an example of a generalised idiopathic epileptic seizure recorded from an adult patient at CUH. The data was recorded using a Viasys Nicolet EEG machine with a sampling frequency of 256 Hz. 35

2.12 EEG sample showing two instances of epileptiform activity recorded from an adolescent patient at CUH. The data was recorded using a Viasys Nicolet EEG machine with a sampling frequency of 256 Hz. 36

2.13 Comparison of EEG recorded using Emotiv EPOC and Viasys Nicolet One EEG machines. Approximately 5 seconds of normal background EEG are displayed for each system. (a) Normal background EEG recorded using Viasys Nicolet EEG machine at a sampling frequency of 250 Hz. (b) Normal background EEG recorded using the Emotiv EPOC EEG device at a sampling frequency of 128 Hz. 44

2.14 EEG channel F8 (blue) and gyroscope x-direction (black) and y-direction (red) signals for artefacts caused by nodding and shaking head. 45

2.15 Illustration of ROC_{sens95} and ROC_{spec90}. The ROC curve is plotted in blue. The horizontal, dashed line represents the lower bound of the ROC_{sens95} area used in the epileptiform activity detection algorithm. The vertical green line represents the lower bound of the ROC_{spec90} area, used in evaluating the neonatal seizure detection algorithm. The dotted red line represents random discrimination in the classification task. 50

3.1 Real-time EEG Analysis for Event Detection (REACT) prototype device. . 56

3.2 Real-time EEG Analysis for Event Detection. This architecture shows three potential methods of incorporating artefact detection into the system. (a) indicates rejection of epochs classified as artefact. (b) indicates classifier fusion of seizure and artefact information. (c) represents artefact annotation, performed separately from the seizure classification task. 57

3.3 Probability density function for line length (L) of background EEG (blue) and head-movement artefact (green). 62

3.4 Probability density function for RMS amplitude of background EEG (blue) and head-movement artefact (green). 63

3.5 Probability density function for activity or variance of background EEG (blue) and head-movement artefact (green). 64

3.6 Probability density function for the mobility of background EEG (blue) and head-movement artefact (green). 64

3.7 Probability density function for the variance of the 1^{st} derivative of background EEG (blue) and head-movement artefact (green). 65

3.8 Probability density function for the variance of the 2^{nd} derivative of background EEG (blue) and head-movement artefact (green). 65

3.9 Probability density function for the kurtosis of background EEG (blue) and head-movement artefact (green). 67

3.10 Probability density function for the non-linear energy of background EEG (blue) and head-movement artefact (green). 68

3.11 Probability density function for the number of zero crossings of the 1^{st} derivative of background EEG (blue) and head-movement artefact (green). . 68

3.12 Probability density function for the number of zero crossings of the 2^{nd} derivative of background EEG (blue) and head-movement artefact (green). . 69

3.13 Probability density function for the autoregressive model fit (1^{st} order) of background EEG (blue) and head-movement artefact (green). 70

3.14 Probability density function for the SEF_{80} of background EEG (blue) and head-movement artefact (green). 71

3.15 Probability density function for the total power in the 0 to 12 Hz range of background EEG (blue) and head-movement artefact (green). 72

3.16 Probability density function for the Shannon entropy of background EEG (blue) and head-movement artefact (green). 74

3.17 Sensitivity, specificity and accuracy for 12-feature, participant dependent LD classifiers. Results are displayed for window lengths of 0.25, 0.5, 0.75, 1, 1.5, 2, 3 and 4 seconds. 83

3.18 Head-movement artefact detection system based on support vector machines classifier. Classification is performed on each EEG channel separately. 88

3.19 ROC areas of participant independent SVM classifiers for a range of window lengths. 89

4.1 Emotiv EPOC headset: Gyroscope axes. 92

4.2 EEG channel F8 (blue) and gyroscope x-direction (black) and y-direction (red) signals for artefacts caused by raising nodding and shaking head. . . . 93

4.3 EEG channel FC5 and all 5 gyroscope channels for artefacts caused by nodding and shaking of the head, in addition to sections of background EEG activity. 93

4.4 EEG electrode F3 and all 5 gyroscope channels for artefacts caused by raising eyebrows and roll head. 94

4.5 EEG electrode F7 and all 5 gyroscope channels for artefacts caused by clenched jaw. 95

4.6 Probability density functions of a selection of gyroscope features corresponding to sections of EEG where head-movement artefact was registered and gyroscope features corresponding to background EEG. The pdfs displayed describe features extracted for window lengths of 1 second. 97

4.7 Probability density functions of a selection of gyroscope features corresponding to sections of EEG where head-movement artefact was registered and gyroscope features corresponding to background EEG. The pdfs displayed describe features extracted for window lengths of 1 second. 98

4.8 Information gain as function of the number of features used, as ranked by the mutual information evaluation function for a window length of 1 second. 104

4.9 System architecture for gyroscope linear discriminant analysis head-movement artefact detector. 106

4.10 LDC results for each of the examined window lengths. 107

4.11 System architecture for gyroscope support vector machine head-movement artefact detector. 108

4.12 Boxplots of SVM ROC areas for each of the examined window lengths for the participant independent experiment. 109

4.13 Boxplot of mean ROC areas for a window lengths of 0.5 and 0.75 seconds for the participant independent experiments. EEG and gyroscope results are shown. 110

5.1 Overview of the feature fusion classifier combination architecture. Feature fusion as illustrated in the diagram is repeated for each EEG channel. . . . 114

5.2 Boxplot of ROC areas for feature fusion classifier with an epoch length of one second. Boxplots of the EEG and gyroscope classifiers are also displayed. Mean ROC areas are shown by the blue line. 115

5.3 Overview of the score fusion classifier combination architecture. The process illustrated in the diagram is repeated for each EEG channel. 116

5.4 Boxplot of ROC areas for score fusion classifiers with an epoch length of one second. Boxplots of the EEG and gyroscope classifiers are also displayed. Median ROC areas are shown within the boxplots by the red lines. Mean ROC areas are shown by the blue line. 119

5.5 Mean ROC plots for score fusion classifier combinations. Mean ROC plots for EEG and gyroscope classifiers are also displayed. 120

5.6 Overview of the decision fusion classifier combination architecture. The decision fusion process highlighted in the diagram is repeated for each channel of EEG. 121

5.7 Boxplot of ROC areas for decision fusion classifiers with an epoch length of one second. Boxplots of the EEG and gyroscope classifiers are also displayed. Mean ROC areas are shown by the blue lines. 122

5.8 Mean ROC plots for decision fusion classifier combinations. Mean ROC plots for EEG and gyroscope classifiers are also displayed. 123

5.9 Boxplot of ROC areas for the best performing feature, score and decision fusion classifiers. Boxplots of the EEG and gyroscope classifiers are also displayed. Mean ROC areas are shown by the blue line. 124

6.1 EEG showing an example of a generalised idiopathic epileptic seizure recorded from an adult patient at CUH. The data was recorded using a Viasys Nicolet EEG machine with a sampling frequency of 256 Hz. 128

6.2 EEG sample showing two instances of epileptiform activity recorded from an adolescent patient at CUH. The data was recorded using a Viasys Nicolet EEG machine with a sampling frequency of 256 Hz. 129

6.3 Breakdown of false detections by underlying cause in epileptiform activity detection classifier. 131

6.4 EEG sample from Patient 2 showing artefacts arising from head movements. The data was collected in the Department of Neurology at CUH using a Viasys Nicolet EEG machine using a sampling frequency of 256 Hz. 132

6.5 EEG sample from Patient 1 showing a number of instances of eye blinks and lateral eye movements. It is evident that the artefacts are most pronounced on frontal electrodes. The data was collected in the Department of Neurology at CUH using a Viasys Nicolet EEG machine using a sampling frequency of 256 Hz. 133

6.6 Overview of the epileptiform activity detection classifier. 134

6.7 Overview of the classifier combination architectures. In block (a), filtering, segmentation and feature extraction is performed on the EEG. In block (b), features are independently presented to three different SVM models. In block (c) the outputs from the SVMs are combined. Note that the threshold (shaded) is applied in block (b) for decision fusion, but in block (c) for score fusion. 135

6.8 Probability density functions of line length for epileptiform EEG (blue) and ocular (green) and movement artefact (red). 136

6.9 Probability density functions of the Power in the 1-3 Hz frequency band for epileptiform EEG (blue) and ocular (green) and movement artefact (red). . 137

6.10 Probability density functions of the number of zero crossings for epileptiform EEG (blue) and ocular (green) and movement artefact (red). 139

6.11 Illustration of classifier output decisions. (a) displays a thirty second section of EEG taken from channel FP2-F8. (b) shows the baseline epileptiform classifier binary decision for the EEG (blue) and annotations (red). (c) displays shows the output decision for the baseline and ocular classifiers deployed together (blue) and the epileptiform annotations (red). 142

6.12 Mean ROC 95 areas (red) and the relative improvement in mean ROC_{sens95} area (blue) for the baseline epileptiform classifier and each score fusion (SF) classifier combination. Random discrimination gives a ROC 95 area of 2.5%. 145

6.13 Boxplots of FDR for the baseline epileptiform classifier and each of the score fusion (SF) classifier architectures. The corresponding mean GDRs are plotted in bold above. 146

6.14 Score (SF) and decision (DF) fusion boxplots of FDR for movement and ocular artefact detection classifiers applied simultaneously in combination with the baseline epileptiform classifier. 147

6.15 Boxplots of FDR for train, test and oracle thresholds. The corresponding mean GDRs are plotted in bold above. 148

7.1 Example of neonatal EEG showing a generalised neonatal seizure. Seizure is present on all channels but is most prominent on channels C4-O2, C3-O1 and C3-T3 (red). 151

7.2 Example of neonatal EEG showing a localised neonatal seizure. Seizure is visible on channels F3-C3, C3-01, Cz-C3 and C3-T3 (red). 152

7.3 Illustration of the ANSeR neonatal seizure detection algorithm developed over the last ten years by the Neonatal Brain Research Group at U.C.C. . . 152

7.4 Example of respiration artefact in EEG recorded from neonatal patients at CUMH using a NicOne EEG machine using a sampling frequency of 250 Hz. Respiration artefact, manifesting as slow waves, on neonatal EEG recording; electrode C4 records the artefact which is then visible on channels F4-C4 and T4-C4, and to a lesser extent on channels C4-P4 and C4-Cz. 154

7.5 Example of respiration artefact in EEG recorded from neonatal patients at CUMH using a NicOne EEG machine using a sampling frequency of 250 Hz. Respiration artefact on neonatal EEG recording; visible on channels F4-C4, C4-O2, T4-C4 and C4-Cz, manifesting as slow and sharp waves. . . 155

7.6 Neonatal seizure detection system with respiratory artefact removal using respiration signal included in the pre-processing stage, as outlined by De Vos et al. (2011). 160

7.7 ROC and ROC_{spec90} areas corresponding to the original neonatal seizure detection performance for two neonatal patients. The blue bars show ROC and ROC_{spec90} areas without explicit artefact removal techniques. The red bars display ROC and ROC_{spec90} areas after removing respiratory artefact utilising SOBI and the respiratory trace as per De Vos et al. (2011). 162

7.8 False detections per hour (FD/h) versus good detection rate (GDR) for Patients 1 and 2, for original neonatal seizure detection system (blue) and the neonatal system with respiratory artefacts removed using SOBI and the respiratory trace (red). 163

7.9 False detections per hour (FD/h) versus threshold for Patients 1 and 2, for original neonatal seizure detection system (blue) and the neonatal system with respiratory artefacts removed using SOBI and the respiratory trace (red). 164

7.10 Illustration of the naming convention for ECG signal with the R-R interval displayed. 165

7.11 Algorithm for obtaining electrocardiogram-derived respiration signal as outlined by Boyle et al. (2009). 165

7.12 Respiratory trace and equivalent EDR signal for 10 seconds of data for Patient 1. 167

7.13 Neonatal seizure detection system with respiratory artefact removal using EDR signal included in the pre-processing stage. 168

7.14 ROC and ROC_{spec90} areas showing original neonatal seizure detection system (blue), and the system with respiratory artefact removed using the respiratory trace (red) and EDR signal (black). 169

7.15 ROC and ROC_{spec90} areas showing original neonatal seizure detection system (blue), and the system with respiratory artefact removed using the electrocardiogram derived respiratory (EDR) signal (black). ROC and ROC_{spec90} areas are reported for four neonatal patients that underwent seizure and whose EEG was contaminated by periods of significant respiratory artefact. 170

7.16 False detections per hour (FD/h) versus good detection rate (GDR) showing original neonatal seizure detection system (blue), and the system with respiratory artefact removed using the electrocardiogram derived respiratory (EDR) signal (black). Results are reported for four neonatal patients that underwent seizure and whose EEG was contaminated by periods of significant respiratory artefact. 172

7.17 Example of motion artefact on the respiratory trace (red) of a neonatal recording at University College London Hospitals. The respiratory trace is relatively artefact-free in the left half of the image; motion artefact causes high amplitude deviations in the right half of the image in both the respiratory trace (red) and EEG signal (blue). 173

References

A. Aarabi, R. Grebe, and F. Wallois. A multistage knowledge-based system for EEG seizure detection in newborn infants. *Clinical Neurophysiology*, 118(12):2781–2797, 2007.

Y.S. Abu-Mostafa, M. Magdon-Ismail, and Hsuan-Tien Lin. *Learning from data*. AML Book, 2012.

V. Afonso, W. Tompkins, T. Nguyen, and S. Luo. ECG beat detection using filter banks. *IEEE Transactions on Biomedical Engineering*, 46(2):192–202, 1999.

A. Al-Ani and M. Deriche. Feature selection using a mutual information based measure. In *Proceedings of the 16th International Conference on Pattern Recognition, Brisbane, Australia*, volume 4, pages 82–85, 2002.

B.Z. Allison, E.W. Wolpaw, and J.R. Wolpaw. Brain–computer interface systems: progress and prospects. *Expert Review of Medical Devices*, 4(4):463–474, 2007.

P. Anderer, S. Roberts, A. Schlögl, G. Gruber, G. Klösch, W. Herrmann, P. Rappelsberger, O. Filz, M. J. Barbanoj, G. Dorffner, and B. Saletu. Artifact processing in computerized analysis of sleep EEG - a review. *Neuropsychobiology*, 40(3):150–157, 1999.

F.S. Bao, D.Y.-C. Lie, and Y. Zhang. A new approach to automated epileptic diagnosis using EEG and probabilistic neural network. In *Proceedings of the 20th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Dayton, Ohio, USA*, volume 2, pages 482–486, 2008.

J.S. Barlow. EMG artifact minimization during clinical EEG recordings by special analog filtering. *Electroencephalography and Clinical Neurophysiology*, 58(2):161–174, 1984.

J.S. Barlow. Automatic elimination of electrode-pop artifacts in EEG's. *IEEE Transactions on Biomedical Engineering*, 33(5):517–521, 1986.

S. Barro, M. Fernandez-Delgado, J.A. Vila-Sobrino, C. V. Regueiro, and E. Sanchez. Classifying multichannel ECG patterns with an adaptive neural network. *Engineering in Medicine and Biology Magazine*, 17(1):45–55, 1998.

A. Bell and T. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444, 1997.

P. Berg and M. Scherg. Dipole modelling of eye activity and its application to the removal of eye artefacts from the EEG and MEG. *Clinical Physics and Physiological Measurement*, 12(A):49–54, 1991.

P. Berg and M. Scherg. A fast method for forward computation of multiple-shell spherical head models. *Electroencephalography and Clinical Neurophysiology*, 90(1):58–64, 1994.

T. Bermudez, D. Lowe, and A. Lamborelle. Schemes for fusion of EEG and ECG towards temporal lobe epilepsy diagnostics. In *Proceedings of the IEEE Engineering in Medicine and Biology Society (EMBC), Vancouver, Canada*, pages 5132–5135, 2007.

J. Bernstein. An overview of MEMS inertial sensing technology. *Zoom Lenses Iii*, 20(2): 14–21, 2003.

C.D. Binnie and H. Stefan. Modern electroencephalography: its role in epilepsy management. *Clinical Neurophysiology*, 110(10):1671–1697, 1999.

B. Boashash, P. Barklem, and M. Keir. Detection of seizure signals in newborns. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Phoenix, Arizona, U.S.A.*, volume 4, pages 2351–2354, 1999.

A.K. Bourke and G.M. Lyons. A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. *Medical Engineering & Physics*, 30(1):84, 2008.

J. Boyle, N. Bidargaddi, A. Sarela, and M. Karunanithi. Automatic detection of respiration rate from ambulatory single-lead ECG. *IEEE Transactions on Information Technology in Biomedicine*, 13(6):890–896, 2009.

C. H. M. Brunia, J. MÃķcks, M. M. C. van der Berg-Lenssen, M. Coelho, M. G. H. Coles, T. Elbert, T. Gasser, G. Gratton, E. C. Ifeachor, B. W. Jervis, W. Lutzenberger, L. Sroka, A. W. van Blokland-Vogelesang, G. van Driel, and J. C. Woestenburg.

Correcting ocular artifacts in the EEG: A comparison of several models. *Journal of Psychophysiology*, 3(1):1–50, 1989.

D. Brunner, R. Vasko, C. Detka, J. Monahan, C. Reynolds III, and D. Kupfer. Muscle artifacts in the sleep EEG: Automated detection and effect on all-night EEG power spectra. *Journal of Sleep Research*, 5(3):155–164, 1996.

A.M.E. Bye, C.A. Cunningham, K.Y. Chee, and D. Flanagan. Outcome of neonates with electrographically identified seizures, or at risk of seizures. *Pediatric Neurology*, 16(3): 225–231, 1997.

J. T. Cacioppo, L. G. Tassinary, and G. Berntson. *Handbook of Psychophysiology*. Cambridge University Press, 2007.

F. Campos Viola, J. Thorne, B. Edmonds, T. Schneider, T. Eichele, and S. Debener. Semi-automatic identification of independent components representing EEG artifact. *Clinical Neurophysiology*, 120(5):868–877, 2009.

J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-gaussian signals. In *IEE Proceedings F: Radar and Signal Processing*, volume 140, pages 362–370, 1993.

A.J. Casson, E. Luna, and E. Rodriguez-Villegas. Performance metrics for the accurate characterisation of interictal spike detection algorithms. *Journal of Neuroscience Methods*, 177(2):479–487, 2009.

P. Celka and P. Colditz. A computer-aided detection of EEG seizures in infants: a singular-spectrum approach and performance comparison. *IEEE Transactions on Biomedical Engineering*, 49(5):455–462, 2002.

C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

G. Chatrian, C. Shaw, and H. Leffman. The significance of periodic lateralized epileptiform discharges in EEG: An electrographic, clinical and pathological study. *Electroencephalography and Clinical Neurophysiology*, 17:77–196, 1964.

R.R. Clancy. Prolonged electroencephalogram monitoring for seizures and their treatment. *Clinics in Perinatology*, 33(3):649–665, 2006.

R.R. Clancy, A. Legido, and D. Lewis. Occult neonatal seizures. *Epilepsia*, 29(3):256–261, 1988.

P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent component analysis and applications*. Elsevier, 2010.

M. Crespo-Garcia, M. Atienza, and J.L. Cantero. Muscle artifact removal from human sleep EEG by using independent component analysis. *Annals of Biomedical Engineering*, 36(3):467–475, 2008.

N. Cristianini and J. Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, 2000.

R. J. Croft and R. J. Barry. Removal of ocular artifact from the EEG: a review. *Clinical Neurophysiology*, 30(1):5–19, 2000.

M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echauz, and B. Litt. Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients. *IEEE Transactions on Biomedical Engineering*, 50(5):603–615, 2003.

N.A.M. de Beer, M. van de Velde, and P.J.M. Cluitmans. Clinical evaluation of a method for automatic detection and removal of artifacts in auditory evoked potential monitoring. *Journal of Clinical Monitoring and Computing*, 11(6):381–391, 1995.

W. De Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel. Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE Transactions on Biomedical Engineering*, 53(12):2583–2587, 2006.

M. De Lucia, J. Fritschy, P. Dayan, and D.S. Holder. A novel method for automated classification of epileptiform activity in the human electroencephalogram-based on independent component analysis. *Medical and Biological Engineering and Computing*, 46 (3):263–272, 2008.

M. De Vos, W. Deburchgraeve, PJ Cherian, V. Matic, RM Swarte, P. Govaert, GH Visser, and S. Van Huffel. Automated artifact removal as preprocessing refines neonatal seizure detection. *Clinical Neurophysiology*, 122(12):2345–2354, 2011.

A. Delorme and S. Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004.

A. Delorme, S. Makeig, and T. Sejnowski. Automatic artifact rejection for EEG data using high-order statistics and independent component analysis. In *Proceedings of the Third International ICA Conference, San Diego,USA*, pages 457–462, 2001.

S. Delsanto, F. Lamberti, and B. Montrucchio. Automatic ocular artifact rejection based on independent component analysis and eyeblink detection. In *Proceedings of the First International IEEE EMBS Conference on Neural Engineering, Capri Island, Italy*, pages 309–312, 2003.

M. Deriche and A. Al-Ani. A new algorithm for EEG feature selection using mutual information. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Salt Lake City, Utah, U.S.A.*, volume 2, pages 1057–1060. IEEE, 2001.

S. Devuyst, T. Dutoit, P. Stenuit, M. Kerkhofs, and E. Stanus. Cancelling ECG artifacts in EEG using a modified independent component analysis approach. *EURASIP Journal on Advances in Signal Processing*, 2008(22):1–13, 2008.

A.A. Dingle, R.D. Jones, G.J. Carroll, and W.R. Fright. A multistage system to detect epileptiform activity in the EEG. *IEEE Transactions on Biomedical Engineering*, 40 (12):1260–1268, 1993.

G. Dornhege, J. del R. Millán, T. Hinterberger, D. McFarland, and K.R. Müller. *Toward brain-computer interfacing*, volume 74. MIT press Cambridge, MA, 2007.

R. Duda, P. Hart, and D. Stork. *Pattern Classification (2nd edition)*. Wiley, 1995.

R.P.W. Duin. The combining classifier: to train or not to train? In *Proceedings of the 16th International Conference on Pattern Recognition, Quebec, Canada*, volume 2, pages 765–770, 2002.

P. Durka and K. Blinowska. Neural networks and wavelet analysis in EEG artefact recognition. In *Proceedings of the Second Conference on Neural Networks and their Applications, Szczyrk, Poland.* http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.53.8225, Accessed on 26/09/2012, 1996.

P.J. Durka, H. Klekowicz, K.J. Blinowska, W. Szelenberger, and S. Niemcewicz. A simple system for detection of EEG artifacts in polysomnographic recordings. *IEEE Transactions on Biomedical Engineering*, 50(4):526–528, 2003.

A.B. Egol and K.K. Guntupalli. Intravenous infusion device artifact in the EEG-confusion in the diagnosis of electrocerebral silence. *Intensive Care Medicine*, 9(1):29–32, 1983.

Emotiv EPOC headset. http://www.emotiv.com/. Accessed on 27/09/2012.

R. Esteller, J. Echauz, T. Tcheng, B. Litt, and B. Pless. Line length: an efficient feature for seizure onset detection. In *Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), Istanbul, Turkey*, pages 1707–1710, 2001.

D. Evans and M. Levene. Neonatal seizures. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 78(1):70–75, 1998.

M. Fatourechi, A. Bashashati, R.K. Ward, G.E. Birch, et al. EMG and EOG artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology*, 118(3):480–494, 2007.

S. Faul, G. Boylan, S. Connolly, W. Marnane, and G. Lightbody. Chaos theory analysis of the newborn EEG: Is it worth the wait. In *IEEE International Workshop on Intelligent Signal Processing, Faro, Portugal*, pages 381–386, 2005a.

S. Faul, L. Marnane, G. Lightbody, G. Boylan, and S.w Connolly. A method for the blind separation of sources for use as the first stage of a neonatal seizure detection system. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Philadelphia, U.S.A.*, volume 5, pages v–409, 2005b.

S. Faul, A. Temko, and W. Marnane. Age-independent seizure detection. In *Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), Minneapolis, U.S.A*, pages 6612–6615, 2009.

S.D. Faul. *Automated neonatal seizure detection.* PhD thesis, Department of Electrical and Electronic Engineering, National University of Ireland, Cork, 2007.

T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

T. Fujioka, N. Mourad, C. He, and L.J. Trainor. Comparison of artifact correction methods for infant EEG applied to extraction of event-related potential signals. *Clinical Neurophysiology*, 122(1):43–51, 2011.

J. Gao, Y. Yang, J. Sun, and G. Yu. Automatic removal of various artifacts from EEG signals using combined methods. *Journal of Clinical Neurophysiology*, 27(5):312, 2010.

T. Gasser, P. Ziegler, and W. F. Gattaz. The deleterious effect of ocular artefacts on the quantitative EEG, and a remedy. *European Archives of Psychiatry and Clinical Neuroscience*, 241(6):352–356, 1992.

T. Gasser, J.C. Schuller, and U.S. Gasser. Correction of muscle artefacts in the EEG power spectrum. *Clinical Neurophysiology*, 116,(9):2044–2050, 2005.

A.S. Gevins, C.L. Yeager, G.M. Zeitlin, S. Ancoli, and M.F. Dedon. On-line computer rejection of EEG artifact. *Electroencephalography and Clinical Neurophysiology*, 42(2): 267–274, 1977.

J.R. Glover, P.Y. Ktonas, M. Shastry, A. Thitai Kumar, and V.M. Muktevi. Methodology and system architecture for automated detection of epileptic seizures in the neonatal EEG. In *Proceedings of the 24th Annual Conference of Engineering in Medicine and Biology Conference (EMBC), Houston, U.S.A.*, volume 1, pages 70–71, 2002.

I.I. Goncharova, D.J. McFarland, T.M. Vaughan, and J.R. Wolpaw. EMG contamination of EEG: spectral and topographical characteristics. *Clinical Neurophysiology*, 114(9): 1580–1593, 2003.

J. Gotman and P. Gloor. Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG. *Electroencephalography and Clinical Neurophysiology*, 41(5):513–529, Nov 1976.

J. Gotman, D.R. Skuce, C.J. Thompson, P. Gloor, J.R. Ives, and W.F. Ray. Clinical applications of spectral analysis and extraction of features from electroencephalograms with slow waves in adult patients. *Electroencephalography and Clinical Neurophysiology*, 35(3):225–235, 1973.

J. Gotman, D. Flanagan, J. Zhang, and B. Rosenblatt. Automatic seizure detection in the newborn: methods and initial evaluation. *Electroencephalography and Clinical Neurophysiology*, 103(3):356–362, 1997.

G. Gratton, M.G.H. Coles, and E. Donchin. A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55(4):468–484, 1983.

B. Greene. *Quantification and classification of electrophysiological markers of seizure in the neonate*. PhD thesis, National University of Ireland, Dublin, 2007.

B.R. Greene, G.B. Boylan, R.B. Reilly, P. de Chazal, and S. Connolly. Combination of EEG and ECG for improved automatic neonatal seizure detection. *Clinical Neurophysiology*, 118(6):1348–1359, 2007.

B.R. Greene, S. Faul, W.P. Marnane, G. Lightbody, I. Korotchikova, and G.B. Boylan. A comparison of quantitative EEG features for neonatal seizure detection. *Clinical Neurophysiology*, 119(6):1248–1261, 2008.

B.R. Greene, A.O. Donovan, R. Romero-Ortuno, L. Cogan, C. Ni Scanaill, and R.A. Kenny. Quantitative falls risk assessment using the timed up and go test. *IEEE Transactions on Biomedical Engineering*, 57(12):2918–2926, 2010.

F. Grouiller, L. Vercueil, A. Krainik, C. Segebarth, P. Kahane, and O. David. A comparative study of different artefact removal algorithms for EEG signals acquired during functional MRI. *Neuroimage*, 38(1):124–37, 2007.

S. Halder, M. Bensch, J. Mellinger, M. Bogdan, A. Kübler, N. Birbaumer, and W. Rosenstiel. Online artifact removal for brain-computer interfaces using support vector machines and blind source separation. *Computational Intelligence and Neuroscience*, Article ID 82069, 10 pages, 2007.

J.J. Halford. Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized ECG interpretation. *Clinical Neurophysiology*, 120(11):1909–1915, 2009.

P.R. Halmos. What does the spectral theorem say? *The American Mathematical Monthly*, 70(3):241–247, 1963.

H. Hassanpour, M. Mesbah, and B. Boashash. Time-frequency feature extraction of newborn EEG seizure using SVD-based techniques. *EURASIP Journal on Applied Signal Processing*, 16:2544–2554, 2004a.

H. Hassanpour, M. Mesbah, and B. Boashash. Time–frequency based newborn EEG seizure detection using low and high frequency signatures. *Physiological Measurement*, 25(4):935, 2004b.

C. He, L. Hotson, and L.J. Trainor. Mismatch responses to pitch changes in early infancy. *Journal of Cognitive Neuroscience*, 19(5):878–892, 2007.

P. He, G. Wilson, and C. Russell. Removal of ocular artifacts from electro-encephalogram by adaptive filtering. *Medical and Biological Engineering and Computing*, 42(3):407–412, 2004.

C.W. Hesse and C.J. James. Tracking and detection of epileptiform activity in multichannel ictal EEG using signal subspace correlation of seizure source scalp topographies. *Medical and Biological Engineering and Computing*, 45(6):909–916, 2007.

J.A. Hirsch and B. Bishop. Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate. *American Journal of Physiology-Heart and Circulatory Physiology*, 241(4):H620–H629, 1981.

B. Hjorth. EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3):306–310, 1970.

G. Holmes. *The Comprehensive Evaluation and Treatment of Epilepsy: A Practical Guide*. Academic Press, 1997.

A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

N. Ille, P. Berg, and M. Scherg. Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology*, 19(2):113, 2002.

K.P. Indiradevi, E. Elias, P.S. Sathidevi, S. Dinesh Nayak, and K. Radhakrishnan. A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram. *Computers in Biology and Medicine*, 38(7):805–816, 2008.

J. Iriarte, E. Urrestarazu, M. Valencia, M. Alegre, A. Malanda, C. Viteri, and J. Artieda. Independent component analysis as a tool to eliminate artifacts in EEG: a quantitative study. *Journal of Clinical Neurophysiology*, 20(4)(4):249–257, 2003.

J.R. Ives and D.L. Schomer. A 6-pole filter for improving the readability of muscle contaminated EEG. *Clinical Neurophysiology*, 69:486–490, 1988.

C. J. James and C. W. Hesse. Independent component analysis for biomedical signals. *Physiological measurement*, 26(1):R15, 2005.

C.J. James and O.J. Gibson. Temporally constrained ICA: an application to artifact rejection in electromagnetic brain signal analysis. *IEEE Transactions on Biomedical Engineering*, 50(9):1108–1116, 2003.

H.H. Jasper. The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10(2):371–375, 1958.

B.W. Jervis, E.C. Ifeachor, and E.M. Allen. The removal of ocular artefacts from the electroencephalogram: a review. *Medical and Biological Engineering and Computing*, 26(1):2–12, 1988.

J.A. Jiang, C.F. Chao, M.J. Chiu, R.G. Lee, C.L. Tseng, and R. Lin. An automatic analysis method for detecting and eliminating ECG artifacts in EEG. *Computers in Biology and Medicine*, 37(11):1660–1671, 2007.

H. Jokeit and M. Schacher. Neuropsychological aspects of type of epilepsy and etiological factors in adults. *Epilepsy & Behavior*, 5:14–20, 2004.

C.A. Joyce, I.F. Gorodnitsky, and M. Kutas. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, 41 (2):313–325, 2004.

T.P. Jung, C. Humphries, T.W. Lee, S. Makeig, M.J. McKeown, V. Iragui, and T.J. Sejnowski. Extended ICA removes artifacts from electroencephalographic recordings. *Advances in Neural Information Processing Systems*, pages 894–900, 1998.

T.P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T.J. Sejnowski. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, 111(10):1745–1758, 2000.

M. Junghöfer, T. Elbert, D.M. Tucker, and B. Rockstroh. Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, 37(4):523–532, 2000.

A. Kapoor, P. Shenoy, and D. Tan. Combining brain computer interfaces with vision for object categorization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage, Alaska, U.S.A.*, pages 1–8, 2008.

N.B. Karayiannis, S. Srinivasan, R. Bhattacharya, M.S. Wise, J.D. Frost Jr, and E.M. Mizrahi. Extraction of motion strength and motor activity signals from video recordings of neonatal seizures. *IEEE Transactions on Medical Imaging*, 20(9):965–980, 2001.

N.B. Karayiannis, A. Mukherjee, J.R. Glover, P.Y. Ktonas, J.D. Frost Jr, R.A. Hrachovy, and E.M. Mizrahi. Detection of pseudosinusoidal epileptic seizure segments in the neonatal EEG by cascading a rule-based algorithm with a neural network. *IEEE Transactions on Biomedical Engineering*, 53(4):633–641, 2006.

S.M. Kay. *Modern spectral estimation*. Pearson Education India, 1988.

D. Kelleher, A. Temko, D. Nash, B. McNamara, and W. Marnane. SVM detection of epileptiform activity in routine EEG. In *Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), Buenos Aires, Argentina*, pages 6369–6372, 2010.

D. Kelleher, A. Temko, S. O'Regan, D. Nash, B. McNamara, D. Costello, and W.P. Marnane. Parallel artefact rejection for epileptiform activity detection in routine EEG. In *Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), Boston, U.S.A.*, pages 7953–7956. IEEE, 2011.

D. Kelleher, A. Temko, S. O'Regan, D. Nash, B. McNamara, D. Costello, and W.P. Marnane. Epileptiform activity detection using support vector machines. *In Review*, 2013.

Z. B. Khaled and G. Farges. First approach for respiratory monitoring by amplitude demodulation of the electrocardiogram. In *Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), Paris, France*, volume 6, pages 2535–2536, 1992.

J. Kittler and F.M. Alkoot. Relationship of sum and vote fusion strategies. In *Multiple Classifier Systems*, pages 339–348. Springer, 2001.

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

M.A. Klados, C. Papadelis, C. Braun, and P.D. Bamidis. REG-ICA: A hybrid methodology combining blind source separation and regression techniques for the rejection of ocular artifacts. *Biomedical Signal Processing and Control*, 6(3):291–300, 2011.

D.W. Klass. The continuing challenge of artifacts in the EEG. *American Journal of EEG Technology*, 35(4):239–269, 1995.

R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Intelligence*, volume 14, pages 1137–1145, 1995.

L. Kuhlmann, A.N. Burkitt, M.J. Cook, K. Fuller, D.B. Grayden, L. Seiderer, and I.M.Y. Mareels. Seizure detection using seizure probability estimation: comparison of features used to detect seizures. *Annals of Biomedical Engineering*, 37(10):2129–2145, 2009.

L.I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley & Sons, 2004.

T.D. Lagerlund, F.W. Sharbrough, and N.E. Busacker. Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. *Journal of Clinical Neurophysiology*, 14(1):73, 1997.

R. Lamothe and G. Stroink. Orthogonal expansions: Their applicability to signal extraction in electrophysiological mapping data. *Medical and Biological Engineering and Computing*, 29(5):522–528, 1991.

P. LeVan, E. Urrestarazu, and J. Gotman. A system for automatic artifact removal in ictal scalp EEG based on independent component analysis and bayesian classification. *Clinical Neurophysiology*, 117(4):912–927, 2006.

O.G. Lins, T.W. Picton, P. Berg, and M. Scherg. Ocular artifacts in EEG and event-related potentials I: Scalp topography. *Brain Topography*, 6(1):51–63, 1993.

A. Liu, JS Hahn, GP Heldt, and RW Coen. Detection of neonatal seizures through computerized EEG analysis. *Electroencephalography and Clinical Neurophysiology*, 82 (1):30–37, 1992.

H. Liu, K.E. Hild, J.B. Gao, D. Erdogmus, J.C. Príncipe, and J.C. Sackellares. Evaluation of a BSS algorithm for artifacts rejection in epileptic seizure detection. In *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), San Fransisco, California, U.S.A.*, volume 1, pages 91–94, 2004.

F. Lopes da Silva. Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalography and Clinical Neurophysiology*, 79(2):81–93, 1991.

F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, B. Arnaldi, et al. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 4, 2007.

S.J. Luck. *An introduction to the event-related potential technique*. MIT Press, 2005.

S. Makeig, A.J. Bell, T.P. Jung, and T.J. Sejnowski. *Advances in Neural Information Processing Systems*, chapter Independent component analysis of electroencephalographic data, pages 145–151. Morgan Kaufmann Publishers, 1996.

M.B. Malarvili and M. Mesbah. Combining newborn EEG and HRV information for automatic seizure detection. In *Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), Vancouver, Canada*, pages 4756–4759, 2008.

D.P. Mandic, D. Obradovic, A. Kuh, T. Adali, U. Trutschell, M. Golz, P. De Wilde, J. Barria, A. Constantinides, and J. Chambers. *Data fusion for modern engineering applications: An overview*. Springer, 2005.

C. L. Mason and L. Tarassenko. Quantitative assessment of respiratory derivation algorithms. In *Proceedings of the IEEE Engineering in Medicine and Biology Society (EMBC), Istanbul, Turkey*, volume 2, pages 1998–2001, 2001.

S. Mathieson. Electrophysiological characteristics of neurological dysfunction in newborns. Neonatal Unit, University College London Hospital, 2012.

R.P. McEvoy, S. Faul, and W.P. Marnane. Ambulatory REACT: Real-time seizure detection with a DSP microprocessor. In *Proceedings of IEEE Engineering in Medicine and Biology Society (EMBC), Buenos Aires, Argentina*, pages 2443–2446, 2010.

D.J. McFarland, A.T. Lefkowicz, and J.R. Wolpaw. Design and operation of an EEG-based brain-computer interface with digital signal processing technology. *Behavior Research Methods*, 29(3):337–345, 1997.

D. McGrath, B.R. Greene, K.J. ODonovan, and B. Caulfield. Gyroscope-based assessment of temporal gait parameters during treadmill walking and running. *Sports Engineering*, pages 1–7, 2012.

D. Meyer, F. Leisch, and K. Hornik. The support vector machine under test. *Neurocomputing*, 55(1):169–186, 2003.

T.M Mitchell. Machine learning, 1997.

J. Mitra, J.R. Glover, P.Y. Ktonas, A.T. Kumar, A. Mukherjee, N.B. Karayiannis, J.D. Frost Jr, R.A. Hrachovy, and E.M. Mizrahi. A multi-stage system for the automated detection of epileptic seizures in neonatal EEG. *Journal of Clinical Neurophysiology*, 26 (4):218, 2009.

E.M. Mizrahi and R.R. Clancy. Neonatal seizures: Early-onset seizure syndromes and their consequences for development. *Mental Retardation and Developmental Disabilities Research Reviews*, 6(4):229–241, 2000.

A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti. ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, 48(2):229–240, 2011.

M.K.I. Molla, M.R. Islam, T. Tanaka, and T.M. Rutkowski. Artifact suppression from EEG signals using data adaptive time domain filtering. *Neurocomputing*, 97:297–308, 2012.

G. Moody, R. Mark, A. Zoccola, and S. Mantero. Derivation of respiratory signals from multi-lead ECGs. *Computers in Cardiology*, 12:113–116, 1985.

N. Mourad, J.P. Reilly, H. De Bruin, G. Hasey, and D. MacCrimmon. A simple and fast algorithm for automatic suppression of high-amplitude artifacts in EEG data. In

Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Honolulu, U.S.A., volume 1, pages I–393, 2007.

D.M. Murray, G.B. Boylan, I. Ali, C.A. Ryan, B.P. Murphy, and S. Connolly. Defining the gap between electrographic seizure burden, clinical expression and staff recognition of neonatal seizures. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 93 (3):F187–F191, 2008.

S. Nagasubramanian, B. Onaral, and R. Clancy. On-line neonatal seizure detection based on multi-scale analysis of EEG using wavelets as a tool. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Chicago, Illinois, U.S.A.*, volume 3, pages 1289–1292, 1997.

H. Nam, T.G. Yim, S.K. Han, J.B. Oh, and S.K. Lee. Independent component analysis of ictal EEG in medial temporal lobe epilepsy. *Epilepsia*, 43(2)(2):160–164, 2002.

K. Nazarpour, H. R. Mohseni, C. Hesse, J. A. Chambers, and S. Sanei. A novel semiblind signal extraction approach for the removal of eye-blink artifact from EEGs. *EURASIP Journal on Advances in Signal Processing*, 2008(1):98, 2008.

H. Nolan, R. Whelan, and R.B. Reilly. FASTER: fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods*, 192(1)(1):152–162, 2010.

S. O'Regan and L. Marnane. Multimodal detection of head-movement artefacts in EEG. *Journal of Neuoscience Methods*, 218(1):110–120, 2013.

S. O'Regan, S. Faul, and W. Marnane. Automatic detection of EEG artefacts arising from head movements. In *Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), Buenos Aires, Argentina*, pages 6353– 6356, 2010a.

S. O'Regan, S. Faul, and W. Marnane. Automatic detection of EEG artefacts arising from head movements using gyroscopes. In *Proceedings of the 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL), Rome, Italy*, pages 1– 5, 2010b.

S. O'Regan, S. Faul, and W. Marnane. Automatic detection of EEG artefacts arising from head movements using EEG and gyroscope signals. *Medical Engineering & Physics*, 35 (7):867–874, 2013a. 10.1016/j.medengphy.2012.08.017.

S. O'Regan, D. Kelleher, A. Temko, B. McNamara, D. Costello, and W.P. Marnane. Robust epileptiform activity detection in the presence of ocular and movement artefacts. *In Review*, 2013b.

S. O'Regan, G. Lightbody, A. Temko, S. Mathieson, G. Boylan, and W.P. Marnane. Respiration artefact removal in automated neonatal seizure detection using blind source separation and an electrocardiogram derived respiratory signal. *In Preparation*, 2013c.

J. Pan and W. J. Tompkins. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, 3(1):230–236, 1985.

S. Park, H. Lee, and S. Choi. ICA+ OPCA for artifact-robust classification of EEG data. In *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing*, pages 585–594, 2003.

Y.T. Peng, C.Y. Lin, M.T. Sun, and C.A. Landis. Multimodality sensor system for long-term sleep quality monitoring. *IEEE Transactions on Biomedical Circuits and Systems*, 1(3):217–227, 2007.

R. Penrose. A generalized inverse for matrices. In *Proceedings of Cambridge Philosophical Soceity*, volume 51, pages 406–413, 1955.

G. Pfurtscheller, J. Kalcher, C. Neuper, D. Flotzinger, and M. Pregenzer. On-line EEG classification during externally-paced hand movements using a neural network-based classifier. *Electroencephalography and Clinical Neurophysiology*, 99(5):416–425, 1996.

G Pfurtscheller, Ch Neuper, D Flotzinger, and M Pregenzer. EEG-based discrimination between imagination of right and left hand movement. *Electroencephalography and Clinical Neurophysiology*, 103(6):642–651, 1997.

T.W. Picton, P. van Roon, M.L. Armilio, P. Berg, N. Ille, and M. Scherg. The correction of ocular artifacts: a topographic perspective. *Clinical Neurophysiology*, 111(1):53–65, 2000.

T.W. Pin, B. Eldridge, and M.P. Galea. A review of developmental outcomes of term infants with post-asphyxia neonatal encephalopathy. *European Journal of Paediatric Neurology*, 13(3):224–234, 2009.

J. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 10(3):61–74, 1999.

R. Polikar, C. Tilley, B. Hillis, and C.M. Clark. Multimodal EEG, MRI and PET data fusion for alzheimers disease diagnosis. In *Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), Buenos Aires, Argentina*, volume 20, page 20, 2010.

M. Qian, M. Aguilar, K.N. Zachery, C. Privitera, S. Klein, T. Carney, and L.W. Nolte. Decision-level fusion of EEG and pupil features for single-trial visual detection analysis. *IEEE Transactions on Biomedical Engineering*, 56(7):1929–1937, 2009.

B. Ramabhadran, J.D. Frost Jr, J.R. Glover, and P.Y. Ktonas. An automated system for epileptogenic focus localization in the electroencephalogram. *Journal of Clinical Neurophysiology*, 16(1):59, 1999.

S.J. Roberts, W. Penny, and I. Rezek. Temporal and spatial complexity measures for electroencephalogram based brain-computer interfacing. *Medical and Biological Engineering and Computing*, 37(1):93–98, 1999.

M. Roessgen, A.M. Zoubir, and B. Boashash. Seizure detection of newborn EEG using a model-based approach. *IEEE Transactions on Biomedical Engineering*, 45(6):673–685, 1998.

S. Romero, MA Mañanas, and MJ Barbanoj. Ocular reduction in EEG signals based on adaptive filtering, regression and blind source separation. *Annals of Biomedical Engineering*, 37(1):176–191, 2009.

G.M. Ronen, S. Penney, and W. Andrews. The epidemiology of clinical neonatal seizures in newfoundland: a population-based study. *The Journal of Pediatrics*, 134(1):71–75, 1999.

A. J. Rowan and E. Tolunsky. *Primer of EEG: With a Mini-atlas*. Butterworth-Heinemann, 2003.

V. Rowland. Cortical steady potential (direct current potential) in reinforcement and learning. *Progress in Physiological Psychology*, 2:1–70, 1968.

M.E. Saab and J. Gotman. A system to detect the onset of epileptic seizures in scalp EEG. *Clinical Neurophysiology*, 116(2):427–442, 2005.

Y. Saeys, I. Inza, and P. Larrañaga. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19):2507–2517, 2007.

Z. Sahul, J. Black, B. Widrow, and C. Guilleminault. EKG artifact cancellation from sleep EEG using adaptive filtering. *Sleep Research*, 24(A):486, 1995.

A. Salarian, H. Russmann, F.J.G. Vingerhoets, C. Dehollain, Y. Blanc, P.R. Burkhard, and K. Aminian. Gait assessment in parkinson's disease: toward an ambulatory system

for long-term monitoring. *IEEE Transactions on Biomedical Engineering*, 51(8):1434–1443, 2004.

S. Sanei and J.A. Chambers. *EEG signal processing*. Wiley-Interscience, 2007.

D. Schachinger, K. Schindler, and T. Kluge. Automatic reduction of artifacts in EEG signals. In *Proceedings of 15th International Conference on Digital Signal Processing*, pages 143–146. IEEE, 2007.

A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller. A fully automated correction method of EOG artifacts in EEG recordings. *Clinical Neurophysiology*, 118(1):98–104, 2007.

B. Schölkopf. The kernel trick for distances. *Advances in Neural Information Processing Systems*, pages 301–307, 2001.

U. Seneviratne, A. Mohamed, M. Cook, and W. D'Souza. The utility of ambulatory electroencephalography in routine clinical practice: A critical review. *Epilepsy Research*, 2013. doi: http://dx.doi.org/10.1016/j.eplepsyres.2013.02.004.

C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1949.

S.Y. Shao, K.Q. Shen, C.J. Ong, E. Wilder-Smith, and X.P. Li. Automatic EEG artifact removal: a weighted support vector machine approach with error correction. *IEEE Transactions on Biomedical Engineering*, 56(2):336–344, 2009.

A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S.T. Treves, and J. Guttag. Patient-specific seizure onset detection. *Epilepsy & Behavior*, 5(4):483–498, 2004.

L. Shoker, S. Sanei, and J. Chambers. Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm. *IEEE Signal Processing Letters*, 12(10):721–724, 2005.

Y. Song. A review of developments of EEG-based automatic medical support systems for epilepsy diagnosis and seizure detection. *Journal of Biomedical Science and Engineering*, 4(12):788–796, 2011.

P. Strobach, K. Abraham-Fuchs, and W. Harer. Event-synchronous cancellation of the heart interference in biomedical signals. *IEEE Transactions on Biomedical Engineering*, 41(4):343–350, 1994.

W. Tatum, B. Dworetzky, and D. Schomer. Artifact and recording concepts in EEG. *Journal of Clinical Neurophysiology*, 28(3):252–263, 2011.

D.M.J. Tax, M. Van Breukelen, R.P.W. Duin, and J. Kittler. Combining multiple classifiers by averaging or by multiplying? *Pattern Recognition*, 33(9):1475–1485, 2000.

A. Temko, E. Thomas, G. Boylan, W. Marnane, and G. Lightbody. An svm-based system and its performance for detection of seizures in neonates. In *Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), Minneapolis, U.S.A.*, pages 2643–2646. IEEE, 2009.

A. Temko, R. McEvoy, D. Dwyer, S. Faul, G. Lightbody, and W. Marnane. REACT: Real-time EEG analysis for seizure detection. In *Proceedings of AMA-IEEE Medical Technology Conference on Individualized Healthcare, Washington, USA*, 2010.

A. Temko, C. Nadeu, W. Marnane, G.B. Boylan, and G. Lightbody. EEG signal description with spectral-envelope-based speech recognition features for detection of neonatal seizures. *IEEE Transactions on Information Technology in Biomedicine*, 15(6):839–847, 2011a.

A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan. EEG-based neonatal seizure detection with support vector machines. *Clinical Neurophysiology*, 122(3):464–473, 2011b.

A. Temko, E. Thomas, W. Marnane, G. Lightbody, and GB Boylan. Performance assessment for EEG-based neonatal seizure detectors. *Clinical Neurophysiology*, 122(3):474, 2011c.

E. Thomas, A. Temko, W. Marnane, G. Boylan, and G. Lightbody. Discriminative and generative classification techniques applied to automated neonatal seizure detection. *IEEE Journal of Biomedical and Health Informatics*, PP(99):1, 2011.

E. M. Thomas. *A Machine Learning Framework for Neonatal Seizure Detection*. PhD thesis, Department of Electrical & Electronic Engineering, National University of Ireland, Cork, 2011.

E.M. Thomas, A. Temko, G. Lightbody, W.P. Marnane, and G.B. Boylan. Gaussian mixture models for classification of neonatal seizures using EEG. *Physiological Measurement*, 31(7):1047–1064, 2010.

O. Tunçel, K. Altun, and B. Barshan. Classifying human leg motions with uniaxial piezoelectric gyroscopes. *Zoom Lenses Iii*, 9(11):8508–8546, 2009.

E. Urrestarazu, J. Iriarte, M. Alegre, M. Valencia, C. Viteri, and J. Artieda. Independent component analysis removing artifacts in ictal recordings. *Epilepsia*, 45(9):1071–1078, 2004.

M. van de Velde, I.R. Ghosh, and P.J. Cluitmans. Context related artefact detection in prolonged EEG recordings. *Computer Methods and Programs in Biomedicine*, 60(3): 183–196, 1999.

S. Vanhatalo, J. Voipio, A. Dewaraja, M.D. Holmes, and J.W. Miller. Topography and elimination of slow EEG responses related to tongue movements. *Neuroimage*, 20(2): 1419–1423, 2003.

V.N. Vapnik. *Estimation of dependences based on empirical data*. Springer-Verlag New York Inc, 2006.

A. Vergult, W. De Clercq, A. Palmini, B. Vanrumste, P. Dupont, S. Van Huffel, and W. Van Paesschen. Improving the interpretation of ictal scalp EEG: BSS–CCA algorithm for muscle artifact removal. *Epilepsia*, 48(5):950–958, 2007.

C. Vidaurre, T. H. Sander, and A. Schlögl. Biosig: The free and open source software library for biomedical signal processing. *Computational Intelligence and Neuroscience*, 2011, 2011.

R.N. Vigário. Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalography and Clinical Neurophysiology*, 103(3):395–404, 1997.

J.J. Volpe. Neonatal seizures: current concepts and revised classification. *Pediatrics*, 84 (3):422–428, 1989.

J.J. Volpe. *Neurology of the Newborn*, volume 1094. Saunders Philadelphia, 2008.

E. Waterhouse. New horizons in ambulatory electroencephalography. *IEEE Engineering in Medicine and Biology Magazine*, 22(3):74–80, 2003.

J.L. Whitton, F. Lue, and H. Moldofsky. A spectral method for removing eye movement artifacts from the EEG. *Electroencephalography and Clinical Neurophysiology*, 44(6): 735–741, 1978.

S.B. Wilson and R. Emerson. Spike detection: a review and comparison of algorithms. *Clinical Neurophysiology*, 113(12):1873–1881, 2002.

I. Winkler, S. Haufe, and M. Tangermann. Automatic classification of artifactual ICA-components for artifact removal in EEG signals. *BEHAVIORAL AND BRAIN FUNCTIONS*, 7(1):30, 2011.

H. Witte, S. Glaser, and M. Rother. New spectral detection and elimination test algorithms of ECG and EOG artefacts in neonatal EEG recordings. *Medical and Biological Engineering and Computing*, 25(2):127–130, 1987.

J. Wolpaw and E.W. Wolpaw. *Brain-Computer Interfaces: Principles and Practice*. OUP USA, 2012.

P. Zarjam, M. Mesbah, and B. Boashash. Detection of newborn EEG seizure using optimal features based on discrete wavelet transform. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Hong Kong*, volume 2, pages 265–268, 2003.

L.I. Zhang, Y. Wang, and H.E. Chuanhong. Online removal of eye blink artifact from scalp EEG using canonical correlation analysis based method. *Journal of Mechanics in Medicine and Biology*, 12(05), 2012.

W. Zhou and J. Gotman. Removal of EMG and ECG artifacts from EEG based on wavelet transform and ICA. In *Proceedings of the IEEE Engineering in Medicine and Biology Conference (EMBC), San Fransisco, U.S.A.*, volume 1, pages 392–395, 2004.

W. Zhou and J. Gotman. Automatic removal of eye movement artifacts from the EEG using ICA and the dipole model. *Progress in Natural Science*, 19(9):1165–1170, 2009.